The Use of an Extended Set of 3D NOESY Experiments for the Structure Determination of Double-Labelled Proteins by NMR

Philip Tobias Robinson

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh
2006

Abstract

Structure determination of proteins by NMR is a well established procedure, widely used in biological and biochemical studies. A method for increasing the efficiency of this technique is proposed whereby a set of 3D NOESY experiments which employ heteronuclear labelling in both indirect dimensions are used to aid assignment of standard 3D ^{15}N- and ^{13}C-edited NOESY spectra.

Using these experiments in concert allows more precise identification of the ^{1}H nuclei involved in NOE interactions and, in doing so, ambiguity in the assignment of NOESY peaks is significantly reduced or eliminated altogether, expediating the process of structure calculation.

The complementary experiments investigated are a 3D ^{13}C,^{15}N-HSQC-NOESY-HSQC and 3D ^{15}N,^{15}N-HSQC-NOESY-HSQC for use with 3D ^{15}N-edited NOESY and 3D ^{13}C,^{13}C-HSQC-NOESY-HSQC for use with 3D ^{13}C-edited NOESY. A methyl selective experiment, 3D ^{13}C,^{13}CH$_3$-HSQC-NOESY-HSQC for use with 3D ^{13}CH$_3$-edited NOESY is also included. Pulse programs for these sequences have been written and implemented. The experiments which gave the greatest benefits to the assignment process are identified.

Semi-automated assignment protocols, developed within the CCPNMR Analysis software package, are developed and used to allow effective manipulation and elucidation of the multiple spectra, streamlining the process. The procedure is tested on the 18 kDa protein, β-Lactoglobulin B and the 15 kDa protein C4BP~1,2. The extent to which assignment ambiguity is reduced relative to a standard assignment method and the effect upon the time requirements for structure determination is investigated. It is shown that, for the sizes of proteins investigated, the time needed for to obtain solution structures is reduced from months to weeks.

During the course of this work, a near-complete resonance assignment of the protein β-Lactoglobulin B is achieved using standard triple resonance experiments for backbone and sidechain assignment and the low pH solution structure of this protein is solved.

Contents

List of Figures

Abbreviations

BLG	β-Lactoglobulin
C4BP	C4b-Binding Protein
CCP	Complement Control Protein
COSY	Correlation Spectroscopy
CSI	Chemical Shift Index
FID	Free Induction Decay
HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PFG	Pulsed Field Gradient
ppm	Parts Per Million
TOCSY	Total Correlation Spectroscopy
RDC	Residual Dipolar Coupling
RMSD	Root Mean Square Deviation

Names and Abbreviations for Amino Acids

Name	Abbreviation	Single Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Chapter 1

Introduction

The process of protein structure determination by NMR (Wuethrich, 1989), unlike direct methods such as X-ray crystallography, is an inherently indirect process. The spectroscopist collects an extensive set of NMR experiments and then uses them, along with a wealth of other, empirical data to determine a family of structures which fit all the experimental and empirical restraints. There are various types of experimental data available to the spectroscopist, but the principal one, on which all NMR-derived structures rely, is the Nuclear Overhauser Effect (NOE). NOESY experiments yield information about the distances which exist between the atoms of the molecule. Observing an NOE between two nuclei indicates that these are within certain distance limits. With a sufficient number of NOE interactions, a structure may be obtained by converting this information into distance limits or restraints for use in structure calculation.

A significant problem in solving protein structures by NMR has been the amount of time that the process requires. This time requirement is a combination of several time-intensive steps in the structure solving process – acquiring spectra, assigning the resonances, deriving inter-proton distances through the assignment of NOESY spectra and the structure calculation process. By far the most significant of these, however, is the assignment of NOESY spectra, and particularly the 3D NOESY spectra commonly used NMR-based protein investigations.

One of the principal difficulties associated with assigning these 3D NOESY spectra is the high degree of degeneracy which exists in the chemical shifts of large proteins. While the use of a third, heteronuclear frequency axis provides much improved signal dispersion, it does not eliminate the problems of degeneracy and, although it is usually possible to unambiguously identify one member of an interacting pair, obtaining the identity of the partner is often non-trivial and there may be several viable candidates. Problems associated with resolution and signal overlap within the spectra further aggravate this process. Consequently, a significant proportion of NOE interactions cannot be attributed to a specific pair of nuclei and the calculation protocol must consider several potential permutations.

The use of 4D spectra should, in theory, remove this problem altogether but such experiments require significantly longer acquisition times and, compared to 3D experiments, have compromised sensitivity and resolution.

In this project, the use of additional 3D NOESY spectra which provide the information which is 'missing' from standard heteronuclear-edited NOESY experiments is investigated. By proceeding in this manner, the spectra benefit from the improved sensitivity and resolution of 3D experiments (relative to 4D) and require no more time to acquire than the equivalent 4D. The most beneficial set of experiments and the manner in which these additional spectra are used, along with the mechanics of their implementation within NMR assignment software, are investigated with the aim of reducing the total time spent on assigning 3D NOESY spectra and, thus, the subsequent structure calculations.

Project Aims

Write and implement complementary NOESY experiments for use with ^{15}N-edited, ^{13}C-edited and $^{13}CH_3$-edited NOESY experiments.

Complete the resonance assignment of β-lactoglobulin B.

Devise a protocol for the use of all the NOESY experiments and implement this protocol within the Analysis (Vranken et al., 2005) software package to allow the experiments to be used together in a simple manner, facilitating quick and efficient assignment of the NOESY spectra.

Determine which experiments are the most useful for obtaining decreased assignment ambiguity and accelerating the structure determination process.

Use the protocol to solve the low pH solution structure of β-lactoglobulin B and quantitatively assess the performance of the protocol relative to the standard method of assigning NOESY spectra.

Perform a further test of the protocol's efficiency by solving the structure of a second protein - C4BP~1,2.

Chapter 2

Background

2.1 Nuclear Magnetic Resonance

NMR relies upon a property of atomic nuclei known as 'spin' which is charcterised by the spin quantum number, I. Nuclei with non-zero values of I possess spin angular momentum. In combination with the charge of the nucleus, this gives rise to a magnetic moment, μ. In the presence of an external magnetic field, B_0, the magnetic moments of such spin-active nuclei will align themselves with respect to the field in discrete, quantised states. The number of states is given by the quantum number, m_I, which can have 2I+1, values (I, I-1, I-2....-I). Values of I are positive multiples of ½ and this is largely an empirical parameter. Conveniently, the nuclei which are of greatest interest to studies of proteins – ^{1}H, ^{13}C and ^{15}N all possess I = ½ and therefore have only two possible spin quantum numbers (+½ and –½) and, thus, two orientations of their spin — corresponding to alignment with or against B_0. These states are termed α and β respectively as shown in Figure 2.1.

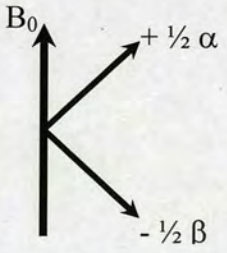


Figure 2.1: Orientations of a spin-1/2 nucleus with respect to a magnetic field, B_0

The magnetic moments are not static, however, and, under the influence of B_0, they precess about the applied field. The frequency of precession, ν (in Hz), is termed the *Larmor* or *resonance* frequency and it is related to the strength of B_0 by equation 2.1:

$$\nu = \gamma B_0 \qquad (2.1)$$

Where $\gamma = \gamma/2\pi$ and γ is a proportionality constant known as the gyromagnetic ratio, which is unique to each nucleus. The energies associated with the α and β states are non-equivalent and it follows that the populations of these two states will therefore be unequal, with a slight excess in the energetically more favourable α state, yielding the population difference necessary to spectroscopy. Transitions between these states can be induced by application of electromagnetic radiation. In order to induce a transition, the applied radiation must have a frequency equal to the resonance frequency. The energy associated with this radiation and, thus, the energy difference, ΔE, between the states is therefore given by equation 2.2:

$$\Delta E = \hbar\gamma B_0 \qquad (2.2)$$

Where $\hbar = h/2\pi$ and h is Planck's constant. Therefore, the larger the magnetic field (and the gyromagnetic ratio), the bigger the splitting in energy. The size of the energy gap falls in the radio frequency (rf) range of the electromagnetic spectrum. Thus, by applying rf radiation, the equilibrium state of a sample can be perturbed. By observing as the system returns to equilibrium, information about the state of the system is obtained. This is the basic principle of NMR.

2.1.1 The Vector Model & Fourier Transformation

For convenience, the example of the ^1H nucleus will be used from now on. From the introductory description of NMR, in a magnetic field, a population difference will exist between the orientations of the magnetic moments of spin-active nuclei. Taken across the whole sample, this slight excess of nuclei in the lower energy level will lead to the

creation of a bulk magnetization vector, M, aligned with B_0 as a result of the contribution of the individual magnetic moments of all the nuclei in the sample. This idea is illustrated in Figure 2.2.

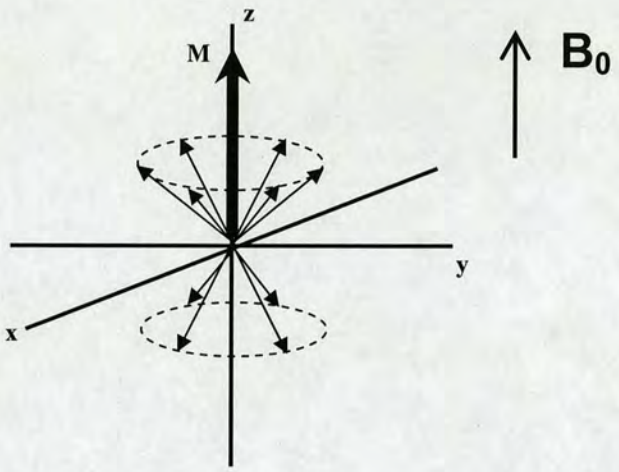


Figure 2.2: Origin of the bulk magnetization vector. An excess of spins (small arrows) in the α state results in the vector, M, aligned with the external field, B_0.

Note that the spins are distributed randomly about the z-axis and, thus, their x-y or transverse components interfere destructively, resulting in zero net magnetization in this plane. The magnetic moment of an rf field, B_1, applied to the sample will interact with this z-magnetization vector, bringing about transformations which correspond to changes in the populations of spins brought about by the transitions induced by the rf field. Ignoring for the moment the behaviour of the individual spins and assuming that B_1 meets the resonance condition outlined previously, applying B_1 in a direction perpendicular to B_0 will cause the vector M to precess about B_1 as illustrated in Figure 2.3. The angle through which the vector moves is determined by the length of time that it is exposed to B_1. For reasons which will be explained later, this is usually of the order of μs and thus these fields are referred to as rf 'pulses'. Typically, angles of 90° ($\pi/2$ pulse) and 180° (π pulse) are used in NMR corresponding to equalisation and inversion of the α and β state populations respectively.

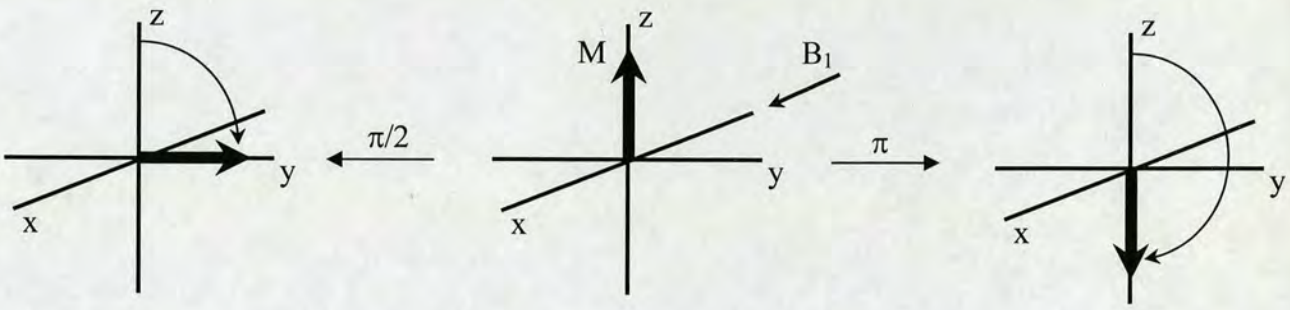


Figure 2.3: Interaction of rf field, B_1 (applied along –x), with magnetization vector, M, showing $90^{\circ}(\pi/2)$ and $180^{\circ}(\pi)$ rotations

In the case of the 90° pulse, it should be noted that this represents something more than simply making the populations equal; a phase coherence has also been imposed upon the spins. Where the distribution of spins about the z-axis had been random, an excess of spins are now aligned along the y axis. It should also be made clear that in these examples a rotating frame of reference is assumed i.e. the 'viewpoint' is moving about the z-axis at the resonance frequency and in the direction of precession.

On removing B_1, M will begin to precess about B_0 at the Larmor frequency, in what is known as free precession, and will return to the z-axis as the populations of spins returns to its equilibrium value. As the transverse magnetization returns to the z-axis, two processes of relaxation are in action. The first mode of relaxation is the loss of phase coherence of the transverse magnetization as a result of the individual spins moving out of phase with one another during free precession. This is spin-spin relaxation and is characterised by the time constant, T_2. The second mode is the restoration of longitudinal (z) magnetization by the various spins of the molecule using the assorted molecular motions of the system to provide or distribute energy as is required for the spins to 'flip' and return to equilibrium. This is spin-lattice relaxation, characterised by the time constant, T_1. The spin-lattice and spin-spin relaxation are brought about by interactions with fluctuating magnetic fields orthogonal to the individual magnetic moments. These fluctuations, coupled with molecular motion act as local pulses and are

the mechanism which eventually leads to the restoration of Boltzmann equilibrium. It is the decaying, transverse magnetization that is detected by the receiver coil of the spectrometer giving rise to the free induction decay or FID. If the magnetization vector is tilted to the y axis, for example, as the spins relax and the magnetization vector precesses about z axis and returns to it, the y component of the magnetization will have the form of a decaying cosine wave as shown in Figure 2.4. The frequency of this wave will be the Larmor frequency of the spin.

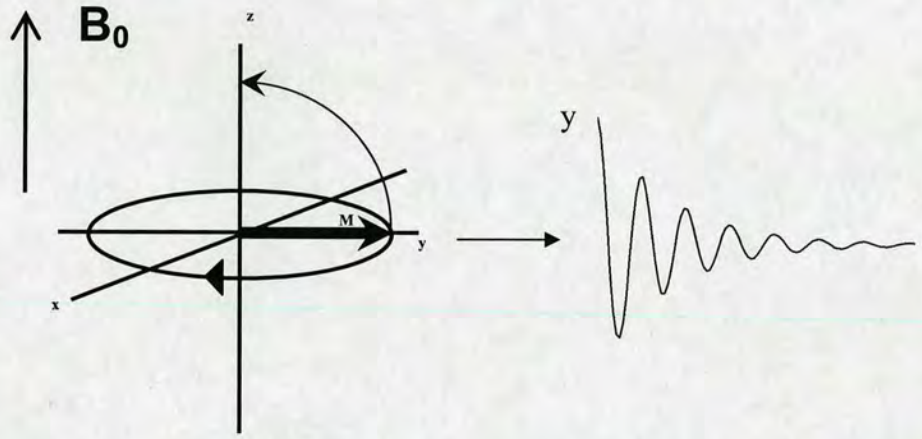


Figure 2.4: Generation of the FID signal – the result of the precession of the transverse magnetization and its gradual return to B_o

The examples used so far are somewhat simplistic in assuming that all spins in a sample possess the same resonance frequency. In reality, this is not the case and individual spins will resonate at different frequencies. The largest contribution to these differences comes from the effect of the magnetic field upon bonding electrons. Under the influence of B_0, the electrons which surround the various nuclei of a molecule are also induced to precess, creating magnetic fields at the nucleus in opposition to B_0. These opposing fields act to 'shield' the nucleus from B_0 and, therefore, the effective magnetic field experienced by a nucleus (and, thus, its resonance frequency) will depend upon the chemical environment of that nucleus. Transverse magnetization created by a $\pi/2$ rotation of M will therefore consist of numerous components, each precessing at a frequency which is characteristic of the chemical environment of the nuclei which gave

rise to it. In practice, resonance frequencies are expressed as a resonance offset, Ω, relative to some reference frequency. This phenomenon is called chemical shift and is given the symbol, δ. It allows the spectroscopist to distinguish between chemically distinct sites within a molecule. This obviously creates a range of rf energies required to match the various different resonance frequencies. By using rf pulses as mentioned previously, the inverse relationship between pulse length and excitation bandwidth is exploited where, for a pulse of length, Δt, there will be a frequency spread, in Hz, of ca. $\pm 1/\Delta t$, and it therefore becomes possible to excite a range of frequencies with a single, short rf pulse.

The full FID of a sample will therefore consist of many individual FID's, one from every environment within the sample and each with an amplitude proportional to the number of spins which exist in that environment. In order to make this primary NMR signal into something easily interpretable, it is necessary to identify the individual frequencies present within the FID together with their amplitudes. This is achieved by the Fourier transform. Put simply, the Fourier transform matches all possible frequencies against the FID to find which ones are present. This yields the simple, 1D NMR spectrum with signals dispersed in terms of their chemical shifts. The intensities of these signals will also be proportional to the amplitude of the FID though it should be noted that this relationship may be distorted during processing of the FID e.g. by applying window functions to the FID.

2.1.2 Scalar Coupling

Nuclei which are separated by a certain number of bonds (usually <3) can 'sense' the state of one another through the bonding electrons. This phenomenon is known as scalar coupling and the nuclei involved are said to be coupled to one another. Taking a simple example of two-spin ½ nuclei, A and X, across the entire sample, all combinations of alignments - $A_\alpha X_\alpha$, $A_\alpha X_\beta$, $A_\beta X_\alpha$, $A_\beta X_\beta$ - will occur. Considering a single A spin,

whatever its orientation, it will 'see' X either aligned with itself or against. Each of these situations will give rise to a further, small change in the magnetic field experienced by A and, thus, the resonance frequency of A will also be altered. If X is aligned *with* the magnetic field, A is shielded more and, thus, will have a lower chemical shift value. If X is aligned *against* the field, the opposite is true (assuming that the coupling is positive in sense). The peak corresponding to spin A will therefore be split into two peaks, each with half the intensity. The same will be true for the signal from X. The magnitude of the splitting depends on the nature and number of bonds separating the interacting nuclei and their orientation in space with respect to each other. The value of this splitting is given the symbol J and will be the same for both spins. A stick representation of such a spectrum is given in Figure 2.5.

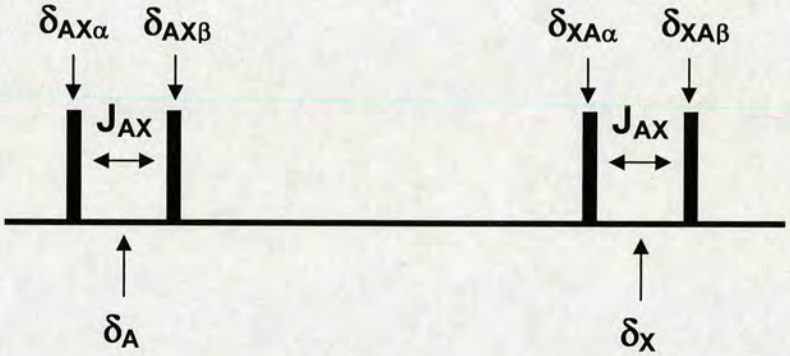


Figure 2.5: Stick representation of the spectrum of two coupled spin ½ nuclei ($\nu_A > \nu_X$, $\Delta\nu >> J_{AX}$)

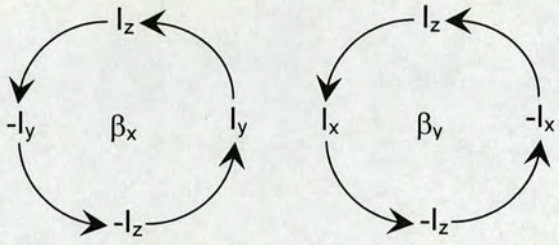
2.1.3 Product Operator Formalism

The vector model is a good visual tool for describing NMR but its usefulness becomes limited when trying to describe complex NMR experiments. The product operator formalism (Sorensen et al., 1983) provides a more convenient method for understanding the workings of NMR, describing the effect of pulses, free precession and coupling on a system in terms of the rotations of operators or products of operators in 'operator space'. A brief summary is given here (adapted from Hore et al., 2000).

x, y and z magnetization vectors of a spin, I, can be represented as the operators I_x, I_y and I_z. The effect of rf pulses, free precession and scalar coupling on these operators can be summarised as follows:

rf pulses:

$$I_x \xrightarrow{\beta_x} I_x$$
$$I_y \xrightarrow{\beta_x} I_y\cos\beta + I_z\sin\beta$$
$$I_z \xrightarrow{\beta_x} I_z\cos\beta - I_y\sin\beta$$

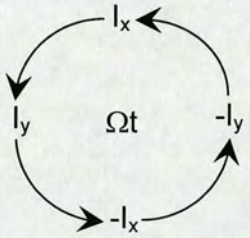


Where β is the flip angle of the pulse and the subscript denotes the phase of the pulse. Thus, as shown, a pulse from x has no effect on x-magnetization. Usually, β is 90° or 180° and so the terms are simplified. Thus, using the example of I_z under the influence of a 90° pulse applied from –x (as illustrated using the vector model in Figure 2.3) gives the following transformation:

$$I_z \xrightarrow{90^\circ_{-x}} I_z\cos 90 + I_y\sin 90 \longrightarrow I_y$$

Free precession:

$$I_x \xrightarrow{\Omega t} I_x\cos\Omega t + I_y\sin\Omega t$$
$$I_y \xrightarrow{\Omega t} I_y\cos\Omega t - I_x\sin\Omega t$$
$$I_z \xrightarrow{\Omega t} I_z$$



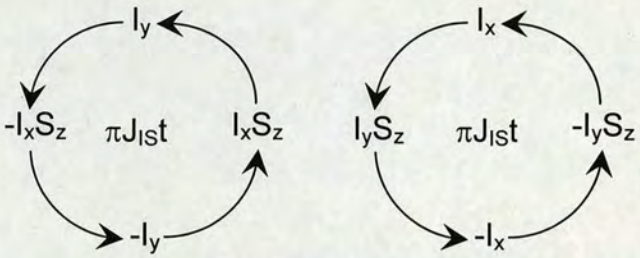
Where Ω is the resonance offset (= 2πν, where ν is the resonance frequency) and t is the duration of the free precession.

Scalar coupling (for two coupled spins, I and S):

$$I_x \xrightarrow{\pi J_{IS}t} I_x\cos\pi J_{XY}t + 2I_yS_z\sin\pi J_{XY}t$$

$$I_y \xrightarrow{\pi J_{IS}t} I_y\cos\pi J_{XY}t - 2I_xS_z\sin\pi J_{XY}t$$

$$I_z \xrightarrow{\pi J_{IS}t} I_z$$



Evolution of magnetization under the influence of J-coupling thus gives new operators which are the products of the operators of the individual spins. These two spin operators transform under the influence of pulses and free precession in the same way as the I_x, I_y and I_z operators.

2.1.4 2D Spectroscopy

2D NMR allows the correlation of nuclei which share some relationship with one another and is a long-standing tool of the NMR spectroscopist (Jeener, 1971; Aue et al., 1976; Kessler et al., 1988). In the most common 2D experiments, nuclei are related either by their scalar couplings (COSY, TOCSY, HSQC) or by their proximity in space via dipolar couplings (NOESY). The whole process can be broadly split into four parts: preparation, evolution, mixing and detection and is easily illustrated by considering a simple, two pulse, COSY experiment illustrated in Figure 2.6.

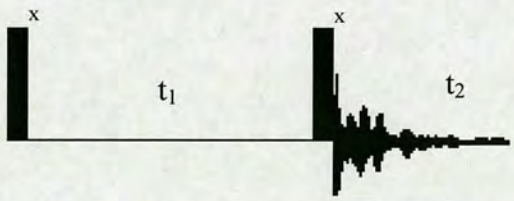


Figure 2.6: A 2D COSY NMR experiment. Black rectangles represent π/2 pulses applied fom the x axis

The description of the first two steps, preparation and evolution, is as follows. The first pulse will tip the magnetization into the xy plane, as has been discussed, where it is then allowed to freely precess at its resonance frequency for a time, t_1. After this period, a

second pulse is applied before detection. Depending on the length of the interval, t_1, some, none, or all of the magnetization will be returned to the z axis before acquisition and, thus, the resultant FID will vary in amplitude according to the length of this period. Considering only a single spin, I, the product operator description of this will be:

$$I_z \xrightarrow{90^\circ_x} -I_y \xrightarrow{\Omega_I t_1} -I_y\cos\Omega_I t_1 + I_x\sin\Omega_I t_1 \xrightarrow{90^\circ_x} -I_z\cos\Omega_I t_1 + I_x\sin\Omega_I t_1$$

The final I_z term is unobservable z-magnetization but it can be seen that the observable I_x term has been modulated by the chemical shift of spin I. Collecting many FID's with gradually incrementing values of t_1, will yield a collection of amplitude modulated spectra. Performing two Fourier transformations, one in the direction of t_1 and one in the direction of t_2 will give a 2D spectrum with a signal appearing at coordinates (δ_I, δ_I). This is known as an autocorrelation or diagonal peak. Figure 2.7 illustrates the process.

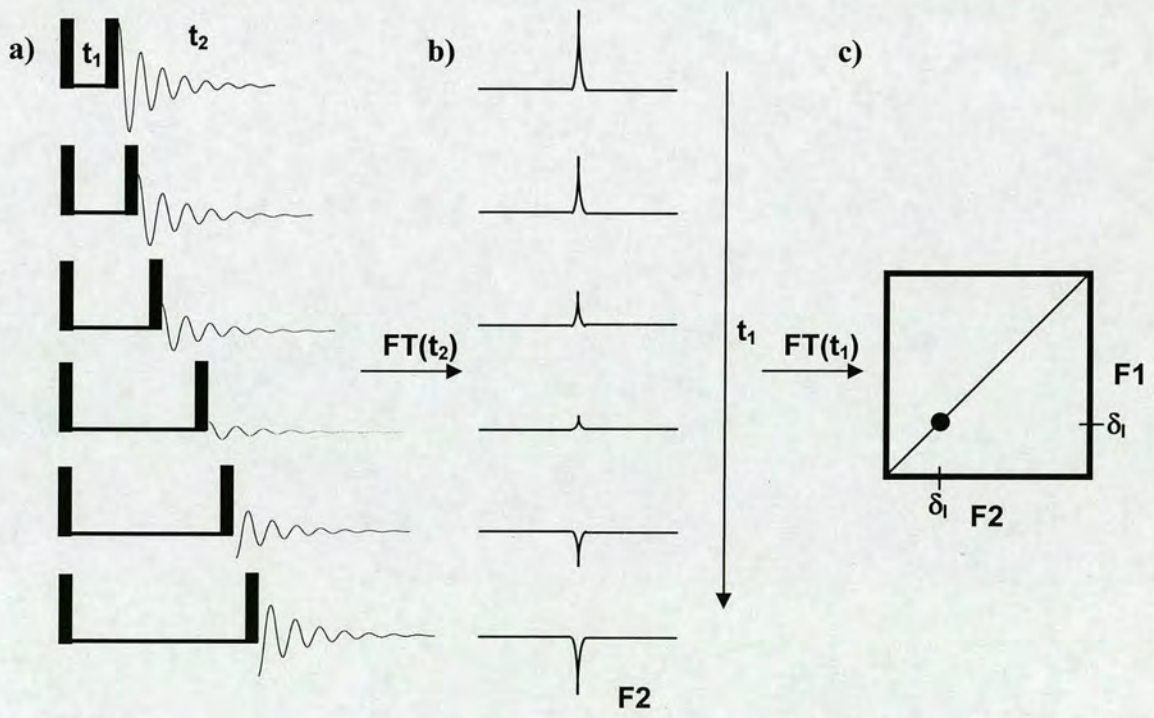


Figure 2.7: Schematic representation of a 2D COSY experiment. a) A series of FIDs are collected by increasing the length of the t_1 interval b) When applied to an isolated spin I, the first Fourier transform yields an array of amplitude-modulated spectra (interferogram) c) The second Fourier transform yields a 2D spectrum. Diagram adapted from Hore, 1995.

This, as it stands is not immensely useful but when this pulse sequence is applied to J-coupled spins, I and S, it results in the transfer of magnetization from spin I to spin S (and vice-versa).

For two coupled nuclei, I and S, The state of this spin system, prior to the final $\pi/2$ pulse is given as:

$$I_z \xrightarrow{90^\circ_x} \xrightarrow{\Omega_I t_1} -I_y\cos\Omega_I t_1 + I_x\sin\Omega_I t_1$$

$$\xrightarrow{\pi J_{IS} t_1} -I_y\cos\Omega_I t_1\cos\pi J_{IS} t_1 + 2I_xS_z\cos\Omega_I t_1\sin\pi J_{IS} t_1$$

$$+ I_x\sin\Omega_I t_1\cos\pi J_{IS} t_1 + 2I_yS_z\sin\Omega_I t_1\sin\pi J_{IS} t_1$$

$$\xrightarrow{90^\circ_x} -I_z\cos\Omega_I t_1\cos\pi J_{IS} t_1 - 2I_xS_y\cos\Omega_I t_1\sin\pi J_{IS} t_1 + I_x\sin\Omega_I t_1\cos\pi J_{IS} t_1$$

$$- 2I_zS_y\sin\Omega_I t_1\sin\pi J_{IS} t_1$$

In this case, only the I_x and I_zS_y terms lead to observable signal. The I_x term is the diagonal peak, as explained above but the I_zS_y term represents magnetization of spin S which will precess at a frequency Ω_S during t_2 but which has been modulated by Ω_I during t_1. Thus, performing two Fourier transformations as before will give a crosspeak with coordinates (δ_I,δ_S). The same will be true for spin S, giving a second diagonal peak and a symmetry-related crosspeak as shown schematically in Figure 2.8.

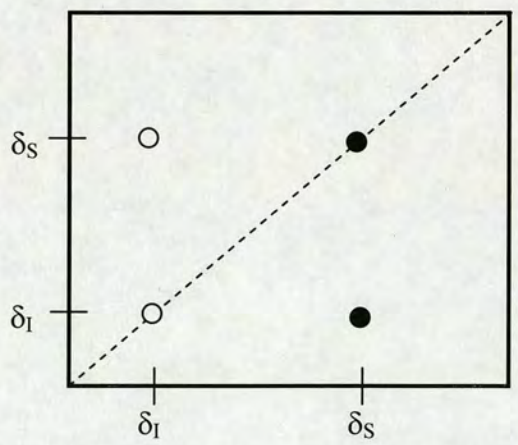


Figure 2.8. Cartoon representation of a 2D spectrum for a two-spin system

In reality, the precise from of these peaks is more complex than shown in Figure 2.8 since the diagonal peaks and crosspeaks are 90° out of phase; the coupling J_{IS} appears in phase and in antiphase for the diagonal peaks and crosspeaks respectively.

In the COSY experiment the preparation consisted of a relaxation period and the first $\pi/2$ pulse, the evolution period was the free precession period and mixing was achieved by the second $\pi/2$ pulse. This principle is the same for all other 2D experiments; the preparation and evolution of magnetization followed by some coherence transfer before, finally, detection.

2.1.5 Nuclear Overhauser Effect

The Nuclear Overhauser Effect or Enhancement (NOE) is a consequence of dipole-dipole interactions between nuclei that are close to one another in space. It is born of relaxation phenomena and results in a change in signal intensity for interacting nuclei. The relaxation of excited spins back to equilibrium has already been discussed but the mechanism of NOE needs some further explanation.

The transfer of energy to or from a spin to cause it to flip is the process by which longitudinal magnetization is recovered. The energy lost by the spins as they return to their equilibrium states is transferred into the various energy modes of the molecular lattice. As has already been noted, however, a spin can only 'flip' if it is exposed to a fluctuating field, fluctuating at the resonance frequency of the spin. It is this field which effects the transfer of energy from spin to lattice or vice-versa. Effectively, a local pulse must be created at the site of an individual spin, causing a spin-flip confined to that spin alone. This local field can be brought about in various ways but it is the interaction of magnetic dipoles or dipole-dipole relaxation which gives rise to NOE.

Two nuclei, whose mutual positions are fixed in the molecular frame and are sufficiently close, will experience a magnetic field due to one another. The orientation of each spin

with respect to B_0 will remain constant but, as the molecule tumbles in solution, the orientations of their magnetic dipoles which they subject one another to, will change over time. As the magnetic dipoles are anisotropic in space, this gives rise to fluctuating magnetic fields. Thus, one spin causes a fluctuating field at the site of the other, effecting spin-flips in the direction perpendicular to the magnetic field fluctuation. These spin flips can occur in an uncorrelated manner (spin I or S flips alone) or in a correlated manner (both spins flip together). Such correlated spin-flips cause so-called cross-relaxation which is at the heart of the NOE since it is this cross-relaxation which allows one spin to affect the populations of another.

Considering the case of two dipolar-coupled nuclei, I and S, shown in Figure 2.9, four possible combinations of states exist for the system.

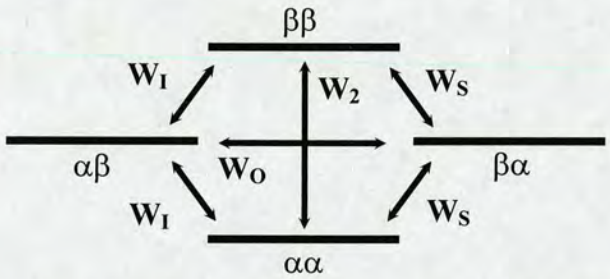


Figure 2.9: The four spin states of a two-spin system. Arrows indicate relaxation transitions arising from dipolar coupling. (Derived from Neuhaus & Williamson, 2000)

Six relaxation pathways are possible for this system with the probability of a transition for the I and S spins *alone* denoted W_I and W_S. W_O and W_2 represent zero and double quantum transitions respectively where both spins flip simultaneously. These cross-relaxation pathways give rise to the NOE since an excited spin relaxing via one of these routes alters the populations and, therefore, intensities of the other spin. The rate of change of the population of a particular state is given by the appropriate combination of probabilities and populations pertaining to that state. So, for the αα state, this description will take the form of equation 2.3:

$$\frac{dn_{\alpha\alpha}}{dt} = -(W_1 + W_S + W_2)n_{\alpha\alpha} + W_1 n_{\beta\alpha} + W_S n_{\alpha\beta} + W_2 n_{\beta\beta} \qquad (2.3)$$

Analogous equations can be used to describe the other states. Since these populations are related to the z-magnetization for I and S, the time dependence of the intensity of the z-magnetization for I and S, ΔI_z and ΔS_z, can be obtained:

$$\frac{d\Delta I_z}{dt} = -\rho_I \Delta I_z - \sigma_{IS} S_z \qquad (2.4)$$

$$\frac{d\Delta S_z}{dt} = -\rho_S \Delta S_z - \sigma_{IS} I_z \qquad (2.5)$$

Where ρ_I and ρ_S are the combined probabilities of I and S spin flips respectively and σ_{IS} is the cross-relaxation rate constant given by the difference between the zero and double quantum transition probabilities (those that produce the NOE). These equations are known as the Solomon equations (Solomon, 1955) and show how z-magnetization evolves during the mixing period of the NOESY experiment. Provided σ_{IS} is non-zero, z-magnetization will be transferred between I and S during the course of the mixing time. The 2D NOESY sequence itself is shown in Figure 2.10.

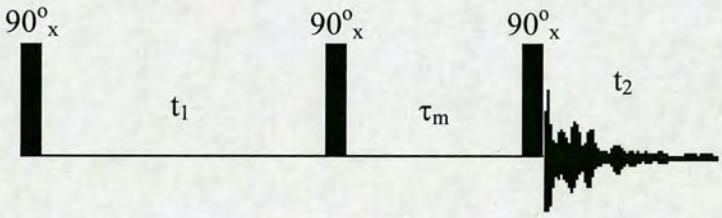


Figure 2.10: The 2D NOESY pulse sequence

As in the 2D COSY experiment discussed before, the preparation and evolution of the magnetization occurs via the action of a $\pi/2$ pulse, followed by a period of free precession, t_1. A second $\pi/2$ pulse then creates the chemical shift modulated (-)z-

magnetization. During the following fixed mixing period, τ_m, cross relaxation occurs and, as a result, magnetization is exchanged between pairs of dipolar coupled nuclei. Assuming, for simplicity, a system with two nuclei which are not J-coupled but are dipolar coupled, and considering only the coherences of interest, the product operator treatment will be:

$$I_z \xrightarrow{90^\circ_x} \xrightarrow{\Omega_I t_1} \xrightarrow{90^\circ_x} -I_z\cos\Omega_I t_1 \xrightarrow{\tau_m} -aI_z\cos\Omega_I t_1 - bS_z\cos\Omega_I t_1$$

$$\xrightarrow{90^\circ_x} aI_y\cos\Omega_I t_1 + bS_y\cos\Omega_I t_1$$

As in the 2D COSY, the final terms represent a diagonal peak (I_y) and a crosspeak (S_y). The coefficients a and b reflect the size of these terms and are dependent upon the length of τ_m and the rate of cross-relaxation.

This technique is immensely useful. Due to the r^{-6} dependence of the NOE, it is only seen for nuclei which are ca. 5Å or less apart and thus provides a great deal of useful information. By showing which atoms are close together, the NOE details how the molecule must be arranged on order for these interactions to be seen. For NMR of proteins, this is the essential aspect of the NOE - the provision of distance restraints.

In folded proteins, the main chains are arranged in various ways, creating the structural features of proteins. In order to solve the protein structure, it is necessary to know more than simply the chemical environment of a spin and its J-coupled partners. As already stated, an NOE crosspeak between two signals gives a maximum distance separation for those two nuclei. Integrating this crosspeak to give its intensity allows a further refinement of that restraint. With sufficient restraints and knowledge of other considerations such as bond lengths, torsion angles etc, a computer program can be used to get a structure for the protein that satisfies all of these restraints. Thus, a full NOESY spectrum should contain much, if not all, of the information required to solve a structure.

A significant problem arises, however, on attempting to extract this information because, as has already been noted, a protein contains many hundreds of protons which will produce a spectrum rich in crosspeaks but complicated by a considerable overlap of signals. Figure 2.11 shows the 2D NOESY spectrum of β-lactoglobulin.

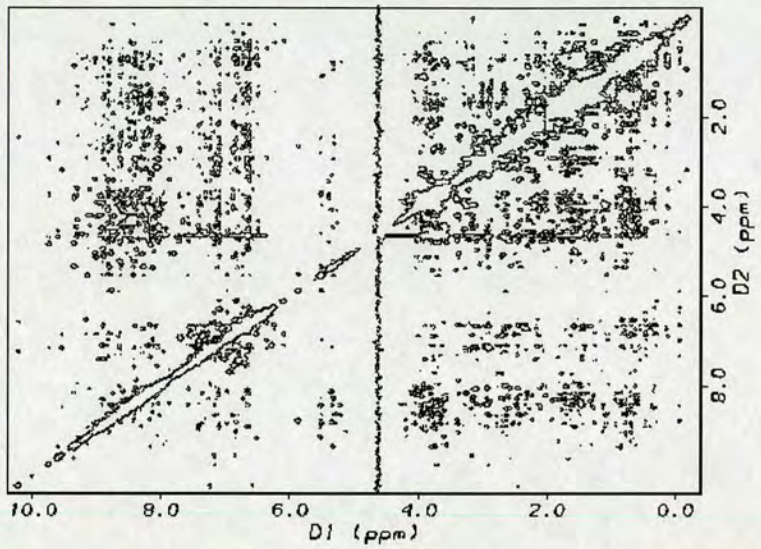


Figure 2.11: The 2D NOESY spectrum of β-Lactoglobulin.

Clearly, the spectrum is immensely complicated and it can be seen that attempting to attribute a particular crosspeak to a pair of signals becomes impossible. For the structure elucidation of larger proteins (ca. >8 kDa), an extra degree of dispersion is required to remove the overlap and allow identification of individual signals and crosspeaks. To this end, isotopic labelling of samples is employed, incorporating ^{15}N (single-labelled) or ^{13}C and ^{15}N (double-labelled) into the protein (Fesik & Zuiderweg, 1988; Marion et al., 1989; Dalvitt et al., 1992).

2.1.6 Heteronuclear correlation

^{13}C and ^{15}N are both NMR active nuclei with I=1/2. When incorporated into proteins they are usually not directly detected but rather used to create an indirectly detected dimension to disperse the signals along another frequency axis.

As described previously, in NMR a correlation can be made between interacting nuclei. COSY and TOCSY use through-bond, scalar coupling and NOESY uses through-space, dipolar coupling. Similarly, it is possible to correlate protons with hetreonuclei by exploiting the coupling that exists between them. For protein NMR, the one-bond heteronuclear coupling between directly bound atoms is the most useful. The method by which magnetization is transferred from protons to their heteronuclei is called INEPT (Insensitive Nuclei Enhanced by Polarisation Transfer, Morris & Freeman, 1979). The INEPT sequence is as shown in Figure 2.12.

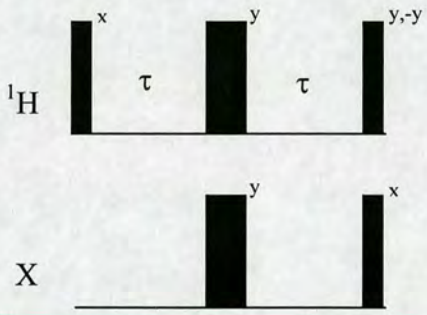


Figure 2.12: The INEPT pulse sequence, X denotes the heteronuclear channel. Thin and thick rectangles represent $\pi/2$ and π pulses respectively.

The three pulses on ^{1}H form a spin-echo during which the proton chemical shifts are refocused and do not have to be considered. Using the product spin operators, the state of a two-spin system H, X, prior to the pair of $\pi/2$ pulses is:

$$H_z + X_z \xrightarrow{(90^\circ_x)_H} -H_y + X_z$$

$$-H_y + X_z \xrightarrow{\pi J_{HX}2\tau,\ (180^\circ_y)_H,\ (180^\circ_y)_X} -H_y\cos\pi J_{HX}2\tau + 2H_xX_z\sin\pi J_{HX}2\tau - X_z$$

If the time, 2τ, for which these evolutions are allowed to occur is chosen to be equal to $1/2J_{HX}$, then the effect is to create a purely antiphase state of the HX coherence, thus:

$$2H_xX_z - X_z$$

Application of the final ^1H $\pi/2$ pulse will then create the zz ordered state, H_zX_z. From here, magnetization transfer to the heteronucleus is achieved by a further $\pi/2$ pulse on the X spin:

$$2H_xX_z - X_z \xrightarrow{(90^\circ_y)_H, (90^\circ_x)_X} 2H_zX_y + X_y$$

Applying rf pulses to the X spin will cause the equilibrium populations of X to be perturbed, creating magnetization which is represented here by the X_y term. To remove this signal, phase cycling or pulsed field gradients can be used, both of which will be discussed later.

When INEPT is part of a multi-dimensional experiment, an incremenTable time delay is used, as explained earlier for homonuclear 2D experiments, to allow labeling of magnetization with the frequency of the heteronucleus. A π pulse is applied to ^1H in the middle of this period to refocus the HX coupling evolution At the end of the labelling period, a reverse INEPT step is used to transfer the magnetization back to protons and to refocus the heteronuclear coupling to yield transverse ^1H magnetization labeled with the frequency of X (Ω_X):

$$H_y \cos\Omega_X t_1$$

This sequence is known as an HSQC (Heteronuclear Single Quantum Correlation) and is shown in Figure 2.13.

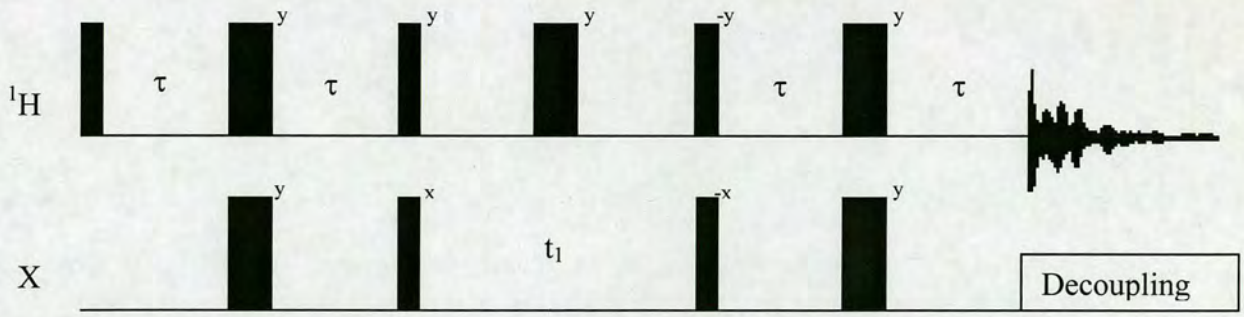


Figure 2.13: The HSQC pulse sequence

This yields a 2D spectrum with the directly detected protons in one dimension and the chemical shift of the heteronucleus in the other.

This building block is used extensively in protein NMR to transfer and label magnetization. Using multiple transfer steps in an experiment allows the correlation of several different frequencies and is the basis of the widely used 3D triple-resonance assignment experiments (Fesik & Zuderweg, 1988; Vis et al., 1994).

Combining polarization transfers and heteronuclear chemical shift labeling steps with a homonuclear correlation in the same experiment, also results in a 3D spectrum. In protein NMR, the combination of a NOESY experiment with HSQC is frequently employed to yield the 3D heteronuclear-edited NOESY which are the mainstay of protein structure determination by NMR (Zuiderweg et al., 1990; Wijmenga & Hilbers, 1990; Ikura et al., 1990; Zhang & Gmeiner, 1996). In such a 3D NOESY experiment the signals of a 2D NOESY spectrum are dispersed along a third axis in terms of a heteronuclear frequency, thus removing much of the signal overlap. Numerous experiments of this sort have been developed and used succesfully in the structure determination of proteins. It is also possible, by a simple extension of this idea, to include a fourth dimension in these experiments so that, for example, all ^{13}C and ^{15}N frequencies can be acquired along with their ^{1}H-^{1}H NOE correlations in a single, 4D experiment (Kay et al., 1990).

2.2 NMR Experimental Tools

The description of experiments given thus far has dealt only with the generation of the desired signal. In reality, performing these experiments requires the implementation of various procedures to suppress these undesired signals, thus ensuring that the complex and elegant sequences deliver the most beneficial results. This is of particular importance in NOESY where the crosspeaks are often very small and so creating artifact-free, well resolved, identifiable signals is paramount.

2.2.1 Coherence selection

In the course of an NMR experiment, as well as creating the desired coherences and achieving their correlation, other, unwanted coherences arise. Magnetization that has not followed the desired pathway must be suppressed or removed.

Phase cycling is a long-standing method of coherence selection (Hoult & Richards, 1975) and involves changing the phase of the rf pulses in the experiment. If this change affects the desired and undesired coherences differently, then phase cycling can be used a means of coherence selection. In the INEPT example described previously (fig 2.12), phase cycling the final ^1H pulse by 180° (y, -y) will lead to the following situations:

$$\text{y:}\quad 2H_xX_z - X_z \xrightarrow{(90^\circ_y)_H} -2H_zX_z - X_z \xrightarrow{(90^\circ_x)_X} 2H_zX_y + X_y$$

$$\text{-y:}\quad 2H_xX_x - X_z \xrightarrow{(90^\circ_{-y})_H} 2H_zX_z - X_z \xrightarrow{(90^\circ_x)_X} -2H_zX_y + X_y$$

Thus, subtracting the signal acquired in two scans will eliminate the X_y term but leave the desired HX coherence.

The second and more recent method of coherence selection is that of pulsed field gradients (PFGs). A PFG is a temporary magnetic field applied, typically for a few milliseconds, with a linear gradient along the length of the sample. The effect of this is to cause identical nuclei in the sample to experience a unique magnetic field depending upon their position in the NMR sample. Applying a PFG to transverse magnetization will therefore dephase it as shown in Figure 2.14 so that on summing over the whole sample no signal is observed. This can be reversed by applying a suiTable rephasing PFG (Vuister et al., 1992; Zhang et al., 1994). Using PFGs in this way to select a particular coherence may result in a loss of half of the signal and is therefore not ideal. Alternatively, the PFG can be used to reject all the unwanted magnetization at a

convenient point in the pulse sequence (Bax & Pochapsky, 1992; Muhandiram et al., 1993). Such a situation exists when the desired magnetization resides in some state that will be unaffected by the action of the PFG e.g. the zz ordered state which is created prior to the polarization transfer pulse of INEPT. Similarly, to remove the X_y term in the INEPT, a $(\pi/2)_X$ pulse at the start of the experiment, followed by a PFG, will destroy the X magnetization and, thus, abolish the X_y term. This can reduce the number of phase cycling steps needed for signal selection. PFGs are also used during the NOE mix time to destroy any transverse magnetization and thus ensure that only the z-component remains.

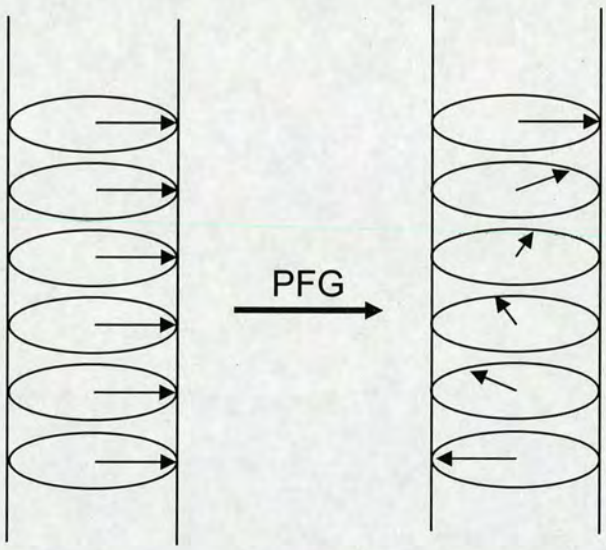


Figure 2.14: Effect on transverse magnetization of a PFG applied along the z-axis

2.2.2 Solvent Suppression

An essential feature of many protein NMR experiments is the suppression of the water signal that is four orders of magnitude more intense than the typical protein resonances

A classical method, which involves the use of PFGs, is known as WATERGATE (Piotto et al., 1992). This uses two PFGs which surround a $\pi/2$-π-$\pi/2$ pulse train, where the $\pi/2$ pulses are water selective. The sequence is shown in Figure 2.15.

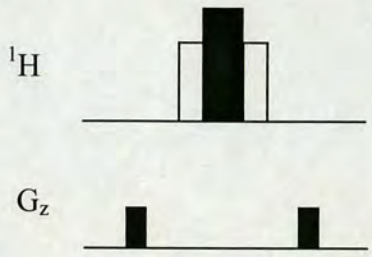


Figure 2.15: The WATERGATE sequence. White bars represent selective water pulses and G_z denotes the gradient channel

Overall, the magnetization of water has experienced a 360° (effectively zero) rotation while the protein proton signals experience only a 180° rotation. Equal PFGs applied around these pulses will therefore dephase the water signal whilst preserving (rephasing) the protein coherences. This technique works well for ^{15}N edited spectra as the signals of interest (amide protons) lie well away from the water resonance and, consequently, will be unaffected by the selective water pulses. For ^{13}C spectra, however, such treatment is unsuiTable as these will contain signals around the water frequency which would be perturbed by the selective pulses. In these cases it is necessary to use other means of solvent suppression such as strong purging gradients combined with phase cycling and/or trim pulses (Majumdar & Zuiderweg, 1993).

2.2.3 Decoupling

By including ^{13}C and ^{15}N in the sample, scalar coupling between protons and these nuclei will exist as explained above. This is useful for correlation but, if left to evolve during labelling periods or acquisition, will lead to a loss in signal intensity and unnecessary signal splitting. Decoupling must be used, then, to remove these effects. During the labelling periods, application of a π pulse in the centre of the period to nuclei other than the one on which the magnetization resides will refocus their couplings and thus ensure that signal loss due to these couplings does not occur. Thus, NH and CH couplings can be removed. This approach can also be used to remove ^{13}C' couplings to

other ^{13}C nuclei as the $^{13}C'$ frequencies are sufficiently separated to allow their selective inversion.

During acquisition, coupling must be removed by other means. Application of a constant irradiating frequency, though simple, is impractical as it requires high power, has a narrow decoupling range and will heat the sample. Instead, decoupling sequences are used. These sequences consist of trains of pulses designed to take the spin which is to be decoupled through a cyclical motion, returning to the starting point at regular intervals. The decoupling sequence, GARP (Shaka et al., 1985), is commonly used and employs a computationally-optimised sequence of composite pulses to manipulate the spins to be decoupled. This allows the power used for the pulses to be smaller than for a constant saturation pulse and gives a wider decoupling range. For >600 MHz NMR spectrometers, however, even this range is insufficient to cover the whole spread of frequencies which are present and adiabatic pulses must be used to give the broad coverage which is required (Kupce & Freeman, 1995; Fu & Bodenhausen, 1995).

2.2.4 Selective Pulses

By reducing the amplitude of rf pulses, the effective fields around which nuclei precess is made offset-dependent. In this case, rather than exciting the entire range of frequencies, only the on-resonance nuclei will experience the nominal flip angles while the off-resonance nuclei progressively experience zero effective rotations.

Such pulses are called selective or 'soft' pulses and allow the selective manipulation of certain spins or groups of spins depending on their chemical shifts. Rectangular selective pulses have unfavourable excitation (and inversion) profiles, producing off-resonance excitation or inversion 'sidebands', but a variety of shaped selective pulses have been developed which have much cleaner profiles. In this project, Q3 and Q5 (Emsley & Bodenhausen, 1990) pulses have been used in triple resonance and NOESY experiments.

2.2.4 Constant Time and Variable Time Chemical Shift Labeling

In the experiments described above, the labeling of magnetization with the chemical shift of a nucleus is achieved by the use an incremented time period. This method of labeling is known as variable time labeling. Labeling can also be achieved via another method known as constant-time labeling. In this approach, the labeling period is not fixed (constant) length, T, in all scans. The varied chemical shift evolution necessary for labeling the magnetization is achieved by applying a 180° pulse to the nuclei being labeled at different times during this evolution period. The scheme is represented in Figure 2.16.

Figure 2.16: Constant-time chemical shift labeling. The first scan has the π pulse in the centre. On subsequent scans, the pulse is shifted toward the start of the sequence.

At first, the 180° pulse sits in the centre of the labeling period and the chemical shift evolution is refocused. On successive scans, moving the 180° pulse (via incrementation of t_1) will allow chemical shift evolution to occur for a period equal to twice the offset of the 180° pulse. For ^{13}C labeling, the length of the period, T, is chosen such that the 1-bond homonuclear coupling, $^1J_{CC}$, which also evolved during this time is refocused at the end of the period.

This approach is useful for labeling ^{13}C resonances since, when using variable time, acquisition times must be kept short to ensure that the $^1J_{CC}$ couplings remain unresolved. Using constant time labeling therefore allows an extended sampling period to be used, increasing the resolution in this dimension. The longer time delay ($1/^1J_{CC}$) means that the scheme is more sensitive to relaxation losses, however, and so it is only suitable for nuclei with favourable relaxation properties e.g. methyl carbons.

2.2.5 Folding

The Nyquist condition states that a wave must be sampled twice in every cycle to correctly define its frequency. The resonance frequencies of nuclei in a sample exist over a certain range and it is generally desirable to detect all the frequencies in this range. The rate at which the FID signal must be sampled to achieve this is determined by the frequency or sweep width (SW) which essentially corresponds to the size or maximum frequency of the spectral 'window'. The time between sampling points is called the dwell time (DW) and, for complex data, is related to the sweep width via equation 2.6:

$$DW = \frac{1}{SW} \qquad (2.6)$$

If the sweep width is chosen such that some resonances lie outside the spectral window, these resonances will still be excited in the experiment but the sampling of the frequencies in indirectly detected dimensions will be insufficient to correctly define their frequency. Signals with a frequency outside the spectral range will be undersampled. Thus the signals will still appear in the spectrum but at artificial frequencies. These signals are said to be folded or aliased. Since it is possible to predict their position, spectral folding can be used to reduce the sweep widths by folding signals into empty regions of the spectrum. This allows increased resolution and reduced acquisition times and is used extensively in ^{15}N and ^{13}C correlated experiments. Folding is not used in the direct dimension where digital filters are used to discard signals outside the spectral window.

2.3 Protein Structure Determination via NMR

Once a protein sample has been prepared, the process of protein structure determination via NMR can be split into four stages – (1) Acquisition of spectra (2) Resonance assignment, (3) NOESY assignment and (4) Structure calculation. These stages will often overlap as it is not necessary that one stage be completed before the next is undertaken. For example, due to the iterative nature of the structure calculation procedure, the last two stages are generally completed together. The whole process, from start to finish, generally has a timescale of several months to a year. This is one of the greatest disadvantages of solving protein structures by NMR and methods to reduce the time spent on each of these stages are continually being investigated. There follows a brief description of some of the more recent and significant advances which have been made in this area.

2.3.1 Acquisition of Spectra

For smaller proteins (ca. < 8kDa) 2D experiments usually provide sufficient information to allow a structure to be solved (Bax, 1989). Such experiments do not require isotopic enrichment and can be acquired in a few days which, when compared to the time requirements of the remaining stages, is a somewhat negligible consideration. Once it becomes necessary, for larger proteins, to use isotopic enrichment and to acquire 3D experiments, however, the time spent on data acquisition becomes much more significant - being on the order of several days for one 3D and, for 4D, a week may be necessary. Various methods have been developed to reduce these lengthy acquisition times and enhance the information content of spectra (Freeman & Kupce, 2003).

The large increase in acquisition time comes from the manner in which these multi-dimensional experiments are conducted – systematically investigating every point in the indirect frequency domains. Thus, every additional dimension will increase the

experiment time significantly. A substantial decrease in experiment time can be achieved by sampling the indirect dimensions simultaneously. Several experiments have been developed in which a third chemical shift is encoded into a multi-dimensional spectrum by labeling two nuclei together in the same dimension (Szyperski et al., 1993a,b; Simorre et al., 1994; Loehr & Rueterjans, 1995), resulting in spectra where the peak positions (in that dimension) are linear combinations of the chemical shifts of the nuclei labeled therein. This effectively results in reducing a nD experiment into n-1D with the associated time gains and the method is known as reduced-dimensionality spectroscopy. The 3D HA CA N HN experiment, for example, is a 3D experiment correlating H^{α} frequencies to C^{α} N^{H} and H^{N} frequencies but with the C^{α} and N^{H} frequencies being labeled together. During t_1, the magnetization is labeled with Hα shifts in the normal way but during t_2 magnetization is labelled with C^{α} shifts and then transferred to N^{H} and labeled with the N^{H} shifts. The result is a spectrum in which, at every H^{α} and H^{N} position, there exist two peaks at $\Omega_{C\alpha}\pm\Omega_{NH}$ as illustrated in Figure 2.17. Such nD spectra require much less acquisition time and afford superior resolution to their n+1D counterparts.

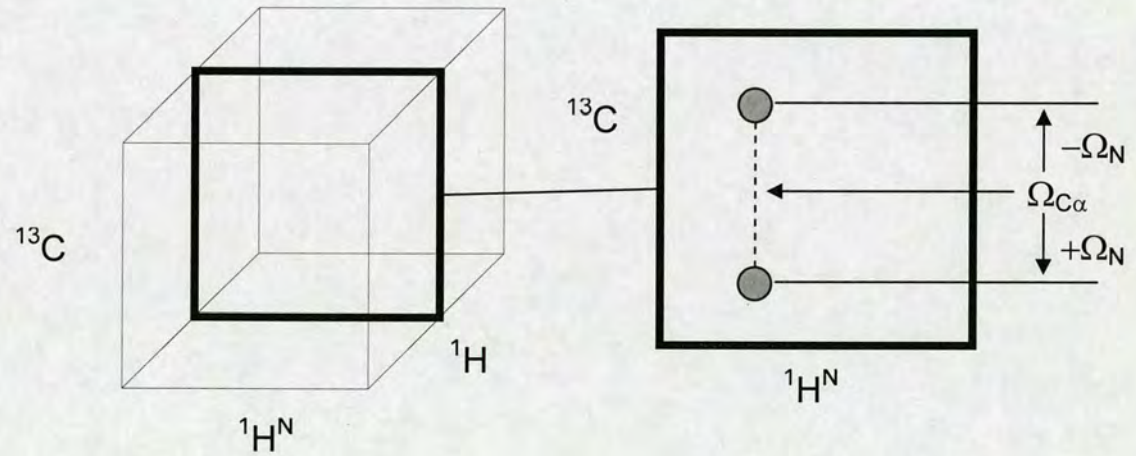


Figure 2.17: Cartoon representation of the reduced-dimensionality HA CA N HN experiment

More recently, an improvement upon this idea known as GFT NMR (Kim & Szyperski, 2002) has been used to combine several indirect dimensions into a single dimension in

such a way as to render the extraction of the individual shifts a relatively simple matter. Obviously, by combining several indirect dimensions, the additional chemical shift information of the indirect dimensions is 'locked' together and must be extracted somehow.

Deconvolution of the signal is achieved by recording the spectrum several times and varying the phases with which the magnetization transfer pulses are applied. After the first evolution period (t_1), magnetization which is transferred to the second evolution period will be modulated by the chemical shift evolution which occurred in t_1, $\cos\Omega t_1$. Similarly, the magnetization proceeding into evolution period, n, will be modulated by $\cos\Omega_1 t_1 \ldots \cos\Omega_{n-1} t_{n-1}$. In the simplest scheme, the incrementation steps are equal as well as simultaneous, giving $\cos\Omega_1 t \ldots \cos\Omega_{n-1} t$. If any of the transfer pulses is shifted by 90°, however, then the modulation function becomes $\sin\Omega_{n-1} t$. Thus, by appropriate phase shifting of these pulses, all combinations of sine and cosine modulation can be obtained. Fourier transform of these signals yields in-phase chemical shift multiplets for the cosine terms and antiphase multiplets for the sine terms. Additional spectra, in which one, some or all but one of the indirect dimensions are 'switched off' by allowing no evolution in these dimensions are also acquired to resolve ambiguities due to overlap. By combining the appropriate spectra, or performing a G-matrix transformation, all n chemical shifts may be obtained. This procedure is illustrated in Figure 2.18. Thus, it becomes possible to acquire previously exotic 5D experiments in a matter of hours. With each additional dimension, the sensitivity will be decreased, however, imposing an eventual sensitivity-limit on the number of dimensions which may be combined.

A similar idea involves simultaneous labeling of ^{15}N and ^{13}C nuclei in a single dimension but with the difference that magnetization is transferred to each nucleus simultaneously, not consecutively and so, rather than combining the chemical shift information, both frequency domains exist on the same axis - effectively obtaining two or more heteronuclear-edited experiments at once (Sørensen, 1990; Farmer II, 1991;

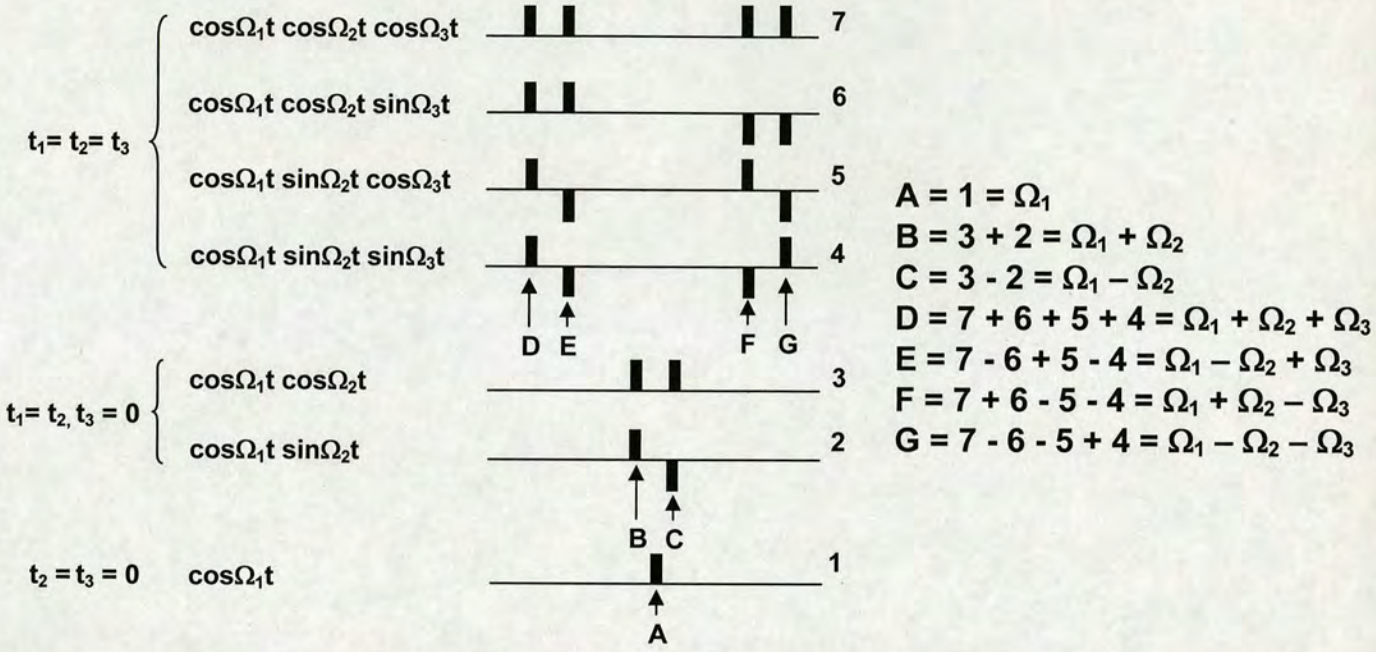


Figure 2.18: Stick figure representation of the multiplets obtained in a GFT NMR experiment in which three dimensions (Ω_1, Ω_2, Ω_3) are recorded simultaneously.

Boelens et al., 1994). Such experiments are known as time-shared or simultaneous acquisition experiments. Applying this to 3D NOESY experiments (Pascal et al., 1994) gives an experiment where the initial magnetization is transferred via NOE to both ^{13}C- and ^{15}N-bound protons, in effect combining a 3D ^{15}N-NOESY and 3D ^{13}C-NOESY spectrum in a single experiment.

This has also been applied to resonance assignment (Vis et al., 1994) and to 4D experiments (Sorensen, 1990; Farmer, 1994). The disadvantage of such methods lies in the differing sizes of e.g. $^1J_{CH}$ and $^1J_{NH}$ couplings. These require different evolution times in the INEPT steps for optimal magnetization transfer, resulting in compromise delay times and, thus, signal loss. Furthermore, the rapidly-evolving ^{13}C-^{13}C couplings which limit ^{13}C labeling periods mean that ^{15}N-labelling periods must be attenuated to some extent when labeling is performed simultaneously.

Hadamard spectroscopy (Kupce & Freeman, 2003) is another method which addresses the fact that comprehensive sampling is not necessary and relies upon the fact that multi-dimensional spectra contain large regions with no information. In Hadamard spectroscopy, specific chemical sites are excited simultaneously (their chemical shifts having been previously obtained) using an array of selective pulses in place of the hard excitation and magnetization transfer pulses normally used in multi-dimensional experiments. The individual responses from each site are encoded using a Hadamard matrix – matrices with orders of 4n, where n is an integer, consisting of plus and minus signs which code for the phase of excitation pulses or the presence of inversion pulses. The overall effect is to change the sign of the final NMR signal for that site. One column of the matrix is required for each individual site and although the number of sites may well be less than the columns of the nearest matrix, the full matrix must be employed. The resulting FIDs are composites of the FIDs from each individual site. In order decode this information, these FIDs are combined in such a way that only the signal from a particular site will add constructively, all others cancelling due to equal contributions of signals with opposite signs. This principle is summarised in Figure 2.19 for a system in which four sites are selectively excited.

	Site			
	A	B	C	D
Scan 1	+	+	+	+
Scan 2	+	–	+	–
Scan 3	+	+	–	–
Scan 4	+	–	–	+

Site B = 1 + 3 – 2 – 4

Figure 2.19: Hadamard matrix for a system in which four sites are excited (A, B, C and D) and the appropriate combination of FIDs which would yield the pure response from site B

Acquisition times for multi-dimensional experiments can be reduced by several orders of magnitude compared to conventional experiments. Of course, the requirement that chemical shifts already be known is something of a disadvantage, particularly for larger

proteins where such information is less easily obtained. Application of this technique is thought to be more useful for binding studies of proteins with known structure and complete resonance assignments where only the signals of interest need be recorded. However, a modification of this idea using band-selective, rather than frequency-selective, pulses has been used (Brutscher, 2004) to simplify 2D spectra used for resonance assignment of a protein by associating the signals of a ^{15}N-HSQC with four different ^{13}CO bands, essentially dispersing the 2D spectrum along a third axis. This idea has been used in combination with reduced-dimensionality experiments to effect rapid resonance assignment of proteins.

Another, conceptually more accessible, mode of reducing acquisition times is presented by the projection-reconstruction method (Kupce & Freeman, 2004). Viewing a 3D spectrum as a cube, projection-reconstruction amounts to projecting the 3D spectrum onto 2D planes using varying angles of incidence, α, with respect to the F_1F_3 plane, or, more simply, taking 'snapshots' from various viewpoints about the cube. The entire spectrum can then be reconstructed by generating the array of peaks which fits all of the projections. Practically, the projections are obtained by simultaneously varying (at different rates) the labeling periods (t_1, t_2) of the indirect dimensions. The time increments, t_1 and t_2, are related to the angle of the projection, α, by the following equations:

$$\mathbf{t_1 = t\cos\alpha}$$

$$\mathbf{t_2 = t\sin\alpha}$$

Thus, when $t_2=0$, $\alpha=0^\circ$ and the projection is the F_1F_3 plane. Similarly, for $t_1=0$, $\alpha=90^\circ$ and the F_2F_3 plane is obtained as shown in Figure 2.20. Other projections at different angles may be obtained by altering t_1 and t_2. Fourier transformation of planes acquired with non-zero t_1 and t_2 yield two projections corresponding to angles of $+\alpha$ and $-\alpha$ with respect to the F_1F_3 plane similar to the sum and difference modulation of shifts in the reduced dimensionality experiments.

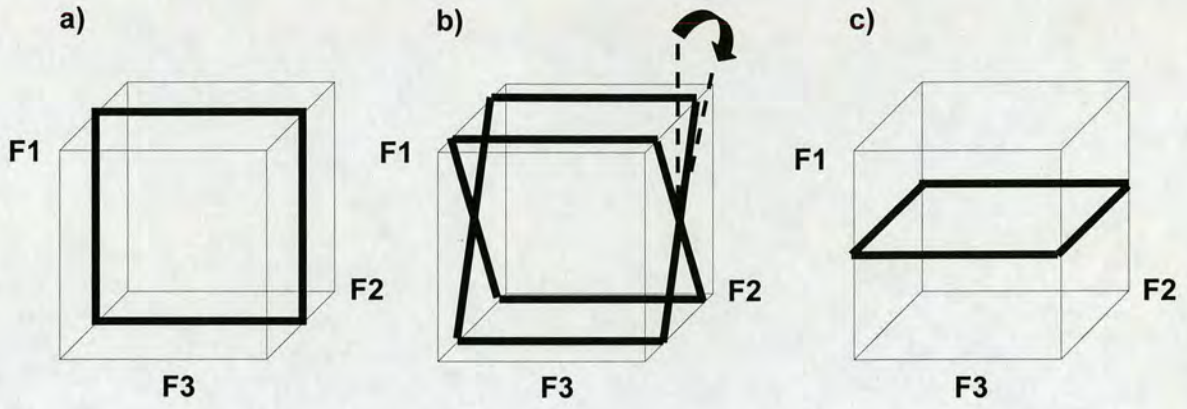


Figure 2.20: Representations of the plane projections obtained using a) $t_2 = 0$, b) $t_1 \neq 0$, $t_2 \neq 0$ and c) $t_2 = 0$

Sufficient projections at appropriate angles will thus yield the entire spectrum in far less time than using conventional methods. The resulting spectrum, however, is identical to that which would have been obtained via linear sampling methods, thus facilitating ease of use.

The greatest reduction in experiment times have been achieved by performing entire experiments in a single scan (Frydman et al., 2002). In this method, rather than exciting the entire sample uniformly and recording chemical shift evolutions over the whole sample by increasing evolution times in consecutive experiments, the single-scan experiment transforms this time separation into a spatial separation by dividing the sample in a number of, effectively, discrete sections, each with a unique t_1 value. This separation is achieved by applying a selective pulse in the presence of strong PFGs which results in only a small section of the sample being excited. Sections of the sample are excited progressively by repeating this initial excitation block whilst incrementing the frequency of the selective pulse. Owing to the time delays between the excitation periods, successive sections evolve with decreasing values of t_1. Thus, the same spins will have evolved for different lengths of time in each slice. Following excitation, a second PFG of opposite polarity is used to remove any precession due to the first PFG so that spins in all slices will have evolved only under their chemical shifts. During

acquisition, the application of another PFG is used to return the spins in all the slices to their original state – essentially undoing the chemical shift evolution which has occurred and 'refocussing' their signals thereby creating an 'echo' in the detected FID. Each nucleus has evolved under its own chemical shift, however, and so the time taken to refocus each spin will differ according to its chemical shift, resulting in 'echoes' occurring at intervals equal to the chemical shift of a particular spin. Thus, the echoes in the FID represent the appearance of the spectrum in the F_1 dimension without any Fourier transformation.

This approach is limited to concentrated samples, however – typically small organic molecules. The sub-millimolar concentrations used in protein NMR are insufficient at present to yield sufficient quality 2D spectra in a single scan.

It should be noted that, while all of the above techniques have been shown to provide significant time gains, their use has been restricted to resonance assignment experiments. Their application to NOESY experiments is much more difficult due to the large dynamic range and high signal density of such experiments.

2.3.2 Resonance Assignment

Depending upon the size of the protein the process of resonance assignment may require weeks to perhaps months of time spent working with spectra. The techniques outlined above have been used to alleviate some of the heavy time demands by providing more information in each experiment to allow faster and more certain assignments to be made. Other developments in this area have pursued the goal of automating the (sequential) assignment process (Moseley & Montelione, 1999). Arguably the largest portion of the time required is spent manually establishing the connectivities between resonances of a 2D root spectrum via some set of 3D experiments and subsequently mapping these assignments onto the protein sequence. A more full description of this process is given in section 3.3. Manual assignment has been the preferred method as it allows the user to

exercise judgement in the identification of spin systems and in making allowances for possible missing peaks or other non-ideal features of the spectrum. Nevertheless, several automated programs have been developed and shown to be effective at rapidly providing accurate assignments. The various programs operate in differing ways but the overall principle is the same in each case – 1) identify the peaks of a spectrum 2) group the peaks into pseudo-residues 3) identify the most likely residue types for a pseudoresidue 4) link pseudoresidues together 5) map the links onto the protein sequence. A comparison of various automated assignment software (Wang et al., 2005) showed the recent MARS (Jung & Zweckstetter, 2004) and CASA programs to be most effective at achieving correct resonance assignments when tested on a range of proteins. These programs have the additional benefit (compared to the others tested) of providing a confidence score for the assignments made. Assignment times are reduced to a matter of minutes but in all cases, complete assignments were not achieved and some erroneous assignments were made. Application of these procedures is therefore semi-automatic, requiring manual input to ensure the process is proceeding effectively and to eliminate potential errors. Often this will result in much more modest time gains. Manual assignment therefore remains the most commonly used method thus far.

While the majority of automated resonance assignment programs focus on the backbone or sequential assignment process, the somewhat more difficult process of automating sidechain assignment has also been reported recently (Masse et al., 2006). Working from an assigned backbone, this program utilises emulation of human logic (via Relative Hypothesis Prioritization) to mimic the detailed decision making processes involved in this procedure. High accuracy of sidechain assignments was achieved for >80% of residues of two medium size (~15 kDa) proteins.

2.3.3 NOESY Assignment & Structure Calculation

Working with NOESY spectra and calculating structures is the most time-intensive stage of protein structure determination via NMR and there has been much work in this area to

improve assignment procedures and develop best practice (Nilges & O'Donoghue, 1998). NOESY spectra contain many thousands of peaks which must be picked and assigned. This process is far from perfect, however and, on completing the NOESY assignment, the first rounds of structure calculation cannot return good quality structures. This is due to a variety of factors such as the potential for erroneous assignments which will distort any structure calculation away from the correct fold; a lack of sufficient restraints to yield a structure; too high a degree of ambiguity in the assignment of peaks; incorrect classification of the peak intensities. Thus, the process must be an iterative one in which the spectroscopist continually returns to the raw data to verify and amend assignments as the calculation protocol identifies potentially incorrect restraints, or to obtain further experimental data if necessary.

For the problem of assignment ambiguity, there exist a variety of programs to aid in converting ambiguous restraints into unambiguous ones. For many of these, the process involves using initial structures calculated using manually assigned peaks. These structures may then be used to edit or filter the ambiguous restraints by removing those assignments which agree poorly with these structures.

The program ARIA (Nilges et al., 1997) is a widely used implementation of such a process and has been incorporated into both CNS and X-PLOR structure generation programs (Brunger et al., 1998; Brunger, 1992). Here, the intensity of an ambiguously assigned peak is assumed to be the superposition of all the various contributions from each possible assignment. The contributions are weighted according to the r^{-6} values obtained from the initial structures. By restricting the number of assignments to those that contribute more than a particular, user-defined fraction of the intensity of a peak, the low level, improbable assignments may be eliminated, leading to more precise structures in the next calculation performed with these filtered restraints.

A similar strategy is employed by the program NOAH (Mumenthaler et al., 1997) which uses initial structures to calculate, for each ambiguously assigned crosspeak, the

percentage of structures in which a particular assignment is violated. The difference in comparison to ARIA being that NOAH creates a restraint for every item in an ambiguous assignment, whereas ARIA provides a single assignment with multiple contributions. Again, those assignments which are violated in more than a set fraction of the structures are discarded. Both of these approaches perform well and have been shown to produce good quality structures from a relatively small number of initial, unambiguous manually assigned NOESY crosspeaks.

A more recent development in this area is the program CANDID (Herrmann et al., 2002) which attempts to remove the need for an initial structure ensemble which the previous programs relied upon. The principal step in achieving this goal is the introduction of some further assessment of the validity of ambiguous NOE assignments prior to performing the initial structure calculations. Ambiguous NOEs are ranked according to their compliance with various criteria including proximity of chemical shifts, symmetry-related crosspeaks and principally, in the first instance, the idea of 'network anchoring'. This idea states that all correct assignments in an NOE dataset should be self-consistent and thus any assignment which is correct should be supported by further assignments within the dataset. For an assignment that links atom 1 to atom 2, for example, this amounts to searching the dataset for further restraints which link atom 1 and atom 2 via some third atom. Restraints which score poorly on this count are more likely to be erroneous assignments and can therefore be discarded. In order to reduce the impact of incorrect restraints which have passed this filtering process, constraint combination is also used whereby long range restraints (those that have the greatest effect upon the structure) are combined together in virtual restraints to increase the likelihood that at least one of the items in each of the resulting ambiguous restraints will be correct. This treatment relies upon the fact that provided at least one of the items of an ambiguous assignment is correct, the structure calculation will not be distorted by other, incorrect contributions. Once a structure has been calculated, this is also included as a means of removing incorrect restraints in subsequent rounds. Using only NOESY peak lists and chemical shift lists as input in structure calculations, CANDID provided >85 % of

assignments in solving the structures of four proteins, the largest of which was a 135 residue protein, calculated using 3D ^{15}N-edited and ^{13}C-edited NOESY spectra along with two 2D NOESY experiments (acquired in H_2O and D_2O). These assignments yielded structures with low rmsd's (<1Å for all backbone atoms). In addition, the use of network anchoring and constraint combination was found to significantly improve the quality of structures at every stage in the process.

A more novel approach to solving protein structures is presented by the CLOUDS protocol (Grishaev & Llinas, 2002a). This treatment eschews the traditional NOE assignment and structure calculation route and provides structures using only NOE peak positions (chemical shifts). The MIDGE (Madrid et al., 1991) protocol is used to specify which protons show an NOE to one another and the distance, derived from crosspeak intensity. This information, along with Van der Waals repulsions as the only other restraint, is then used to generate a proton distribution or 'cloud' via simulated annealing of a proton gas. Numerous calculations are performed to generate a family of clouds, similar to the structure ensembles provided by traditional NMR structure methods. Importantly, the result is a proton-only description of the molecule, not dissimilar to the electron density maps of heavy atoms obtained in X-ray crystallography. The benefits of this procedure are that it requires no assignments, eliminating a large, time-consuming part of the calculation process and thus very quickly yields information on the protein surface. A protocol for transforming this proton density map into a full molecular structure has also been designed and successfully implemented (Grishaev & Llinas, 2002b) giving final structures which show good agreement with those previously calculated via X-ray and NMR.

2.3.4 Using Complementary Experiments to Reduce Ambiguity

The problem of ambiguity is so common because, in any NOE interaction, a 3D experiment can only provide three out of four relevant chemical shifts. Ambiguity can therefore be reduced by acquiring 4D NOESY spectra. In 4D NOESY experiments,

illustrated schematically in Figure 2.21, resonance frequencies of all four atoms of interest are correlated with one another in the spectrum – the chemical shift of both interacting protons and their heteronuclei - resulting in a much more accurate assignment (Grzesiek et al., 1995; Muhandiram et al., 1993; Morshauser & Zuiderweg, 1999; Vuister et al., 1993). 4D spectra require much more time to acquire, however, and, for larger proteins (>15kDa) the sensitivity of these experiments is significantly compromised due to the relaxation properties of such proteins (Kay et al., 1992).

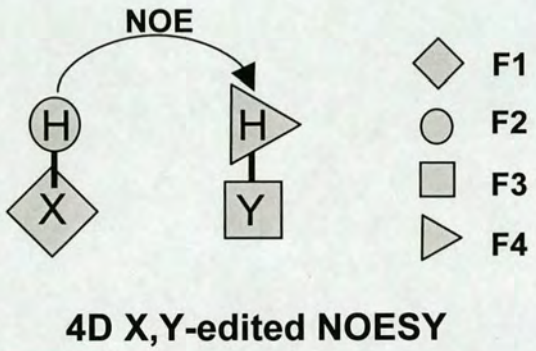


Figure 2.21: Schematic representation of a 4D NOESY experiment

This leads to a loss of the weaker NOE's, which are so important in delivering accurate, good quality structures. Limited sampling of the indirectly detected dimensions of 4D experiments will also decrease the resolution and, thus, increase the ambiguity of the assignments. An alternative to this approach is to split the 4D information into two 3D experiments – a traditional, heteronuclear-edited NOESY spectrum and another spectrum in which the chemical shift of the heteronucleus (which was not labeled in the oroginal experiment) is labeled in t_1. This idea has found several implementations thus far (Frienkel et al., 1990; Diercks et al., 1999; Jerala et al., 1995; Uhrin et al., 2000; Xia et al., 2003) and is illustrated schematically in Figure 2.22.

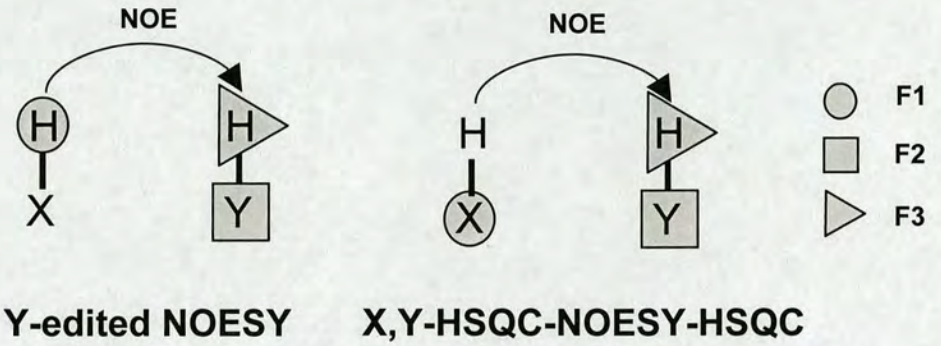


Figure 2.22: Schematic representations of a standard 3D edited NOESY and the associated complementary experiment

Thus it is possible to obtain all four chemical shifts in two experiments, maximising the sensitivity and resolution in each experiment. It is recognized that, compared to a 4D experiment, this approach still leaves room for ambiguity as the link between the F_1 ^1H and X resonances is not firm. It is the focus of this project to establish whether, despite this practical weakness, this method can provide qualitatively faster progress toward protein structures. Furthermore, it is also the intention to conduct a more comprehensive investigation of various 'standard' and complementary experiments to discern those combinations which prove most useful for increasing the efficiency of protein structure determination by NMR and to provide a quantitative appraisal of the differences between the two methods.

To this end, a full description of the two approaches – the standard method (referred to as Method 1) and the new method (referred to as Method 2) is given below.

Method 1

Assigning a 3D NOESY spectrum consists of picking the peaks and then, for a particular peak, assigning a nucleus to each of the three dimensions of that peak. The assignment possibilities in each dimension are determined by the chemical shift list of the protein

which is obtained via the process of resonance assignment (see 3.3). Since a peak will have a particular ppm value associated with each of its dimensions, items from the chemical shift list with chemical shift values within a predefined range may be considered as valid assignment possibilities for that dimension. Ideally, there will be only one possibility for every dimension of every peak. This is seldom the case, however, owing to the high degree of degeneracy which exists in the chemical shifts of such large molecules.

In general, assigning two of the dimensions is a relatively trivial matter owing to the fact that the nuclei found in these dimensions are bonded to one another e.g the $^{1}H^{N}$ and $^{15}N^{H}$ dimensions in a 3D ^{15}N-edited-NOESY. Thus, even though more than one assignment possibility exists for one of these dimensions, only those possibilities which correspond to bonded nuclei can be considered valid, greatly reducing the number of possibilities. More usually, the assignment of 3D NOESY spectra will proceed with an assigned 2D spectrum as its root e.g. a ^{15}N-^{1}H-HSQC. Every peak in the strip of the 3D spectrum extending from a particular peak in the 2D spectrum can then be assumed to have the same assignment in the dimensions it shares with the 2D spectrum. The assignment possibilities in the remaining third dimension are determined by proximity of chemical shift but, in this case, with no information about the bound heteronucleus. Thus, while it may be the case that only a single assignment possibility is obtained, depending upon the size of the protein, it is more likely that there will be a number of possibilities, none of which can be regarded as more or less valid than the others. This difficulty can be addressed by the process of searching for symmetry related peaks. As noted for the 2D NOESY, for any NOE interaction, there will exist two crosspeaks due to the fact that both protons which share the interaction will receive magnetization from each other. For any potential interaction indicated by the assignment possibilities for a peak, that assignment can be confirmed by the existence of a symmetry related peak. Any assignment possibility which does not have a supporting symmetry related peak may be considered an invalid possibility and therefore disregarded. Of course, when working with a heteronuclear edited NOESY, if a crosspeak arises due to an interaction with a

proton bound to a different type of heteronucleus, the symmetry peak will not exist in that spectrum. For example, in the ^{15}N-edited NOESY spectrum, the symmetry-related peaks for $^{1}H^{N}$-$^{1}H^{C}$ NOEs will be found in the ^{13}C-edited NOESY spectrum. This process is illustrated in Figure 2.23.

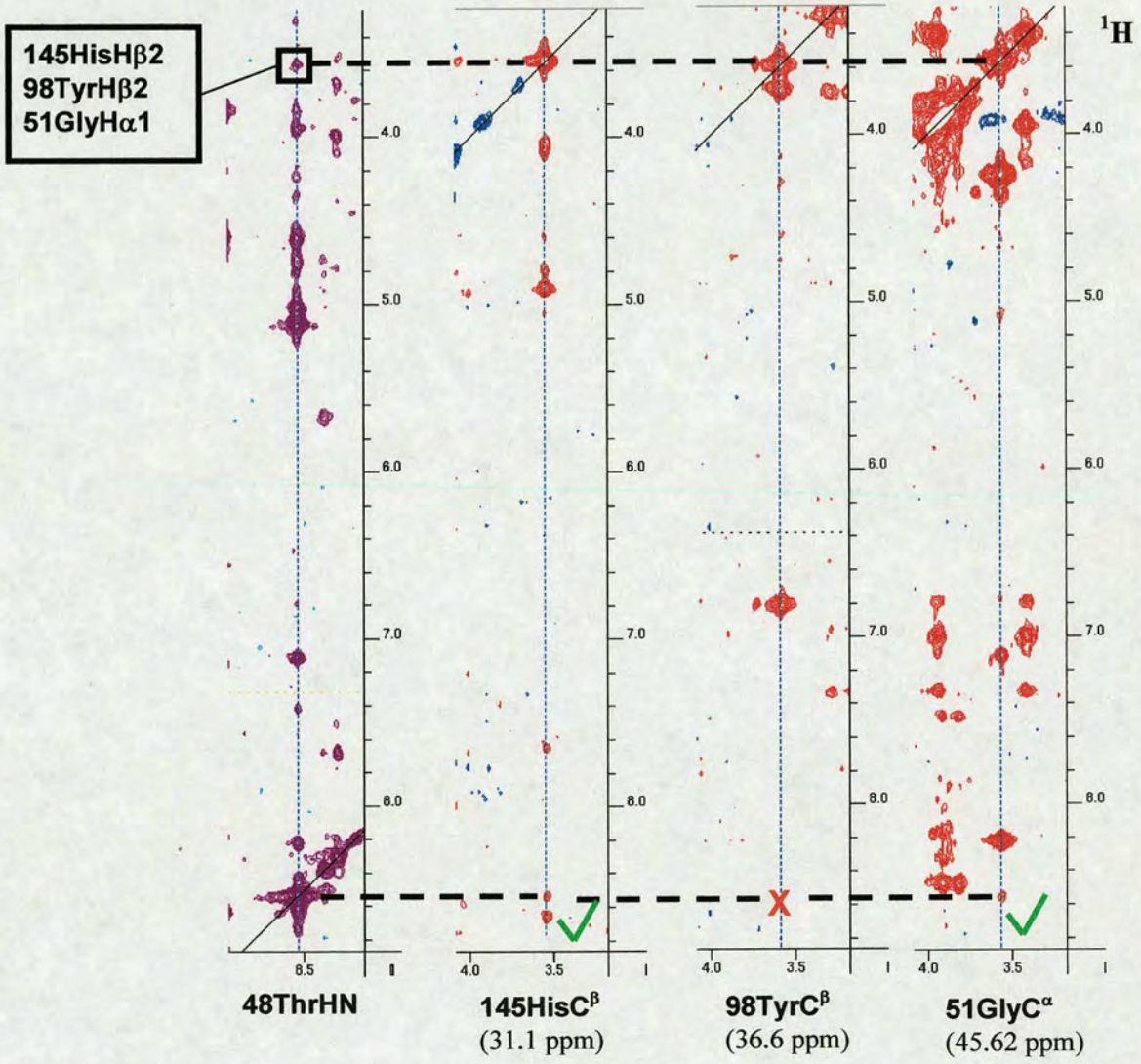


Figure 2.23: Looking for symmetry related peaks. ^{15}N-edited NOESY strip for Thr 48 of β-lactoglobulin B (purple), with strips of the ^{13}C-edited NOESY spectrum at the F_2 positions indicated. Each ^{13}C-edited NOESY strip is examined to see whether a symmetry peak exists. In this case, there are potential symmetry peaks for two of the three assignment possibilities of the boxed ^{15}N-edited NOESY peak.

Thus, the method may be summarised as follows

1 Obtain assignment possibilities
2 Look for symmetry peaks
3 Assign / do not assign

While this certainly reduces the degree of ambiguity, in many cases resulting in an unambiguous assignment, it may still be the case that even after searching for symmetry related peaks, there will remain several valid assignment options, as is the case in the example given in Figure 2.23, where there are potential symmetry related peaks for two of the three of the assignment possibilities for the peak in question. While it may be the case that both of these nuclei share an NOE with the nuclei in question, it may also be the case that one of the symmetry peaks is in fact due to some other NOE interaction and has been mistakenly identified due to the chemical shift degeneracy mentioned previously. It is also possible that the region in which a symmetry related peak should exist will be very crowded or otherwise non-ideal, making the identification of a symmetry related peak impossible.

Method 2

In an effort to alleviate this problem, the use of additional NOESY experiments can be employed to further reduce the number of ambiguous assignments. As described above, these experiments label the heteronucleus of the NOE partner in t_1. Examining equivalent strips of the 'standard' NOESY experiments and the complementary NOESY experiments, gives a further measure of the validity of a particular assignment. This can be used to 'filter' assignment possibilities prior to looking for symmetry related peaks. This idea is illustrated in Figure 2.24 using the same peak as in the previous Figure 2.23. Whereas checking for symmetry related peaks did not produce an ambiguous assignment for the peak under examination, using the complementary NOESY experiment shows that only one of the assignment possibilities is valid and the others may be discarded

thus producing an unambiguous assignment for that peak. In this case, looking for symmetry peaks could be deemed unnecessary, however, this process is included in the method in all instances to provide confirmation of those assignments which are obtained and, as will be discussed later (section 5.2) to reduce the potential for erroneous assignments.

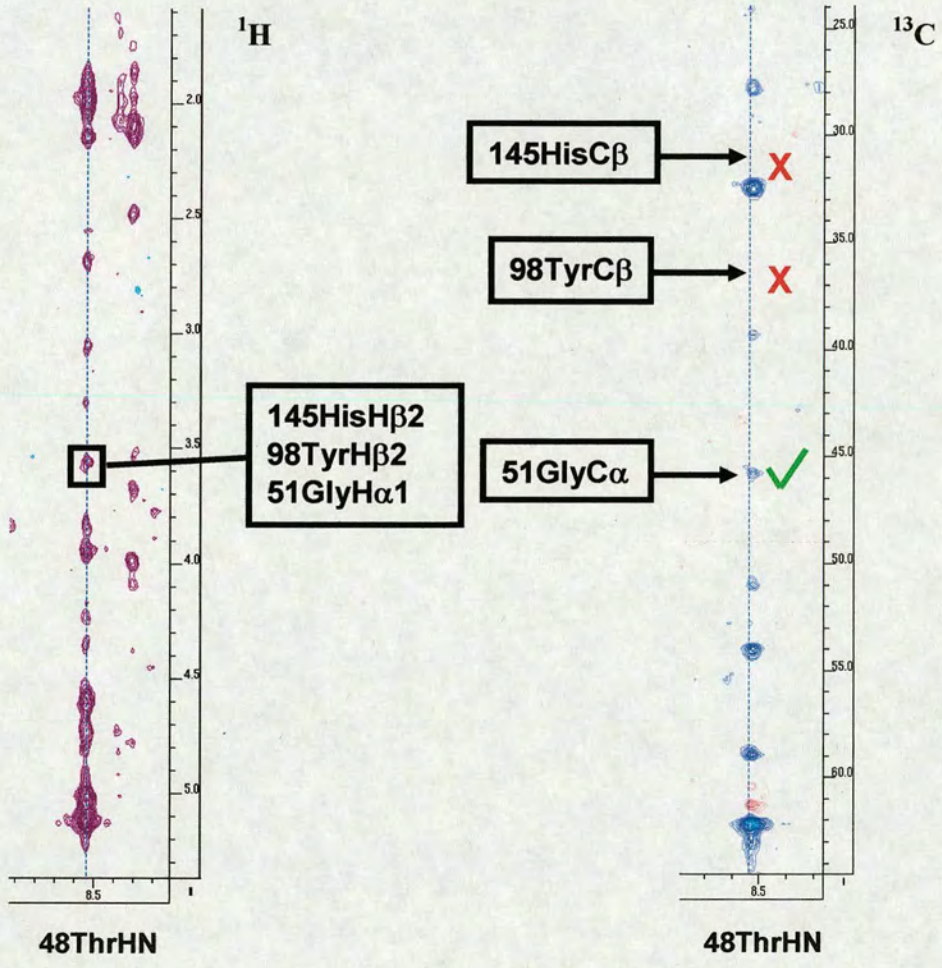


Figure 2.24: ^{15}N-edited NOESY strip with the equivalent strip of the complementary ^{13}C,^{15}N-HSQC-NOESY-HSQC experiment.

Chapter 3

Materials & Methods

3.1 Samples

3.1.1 β-Lactoglobulin B

A ^{13}C,^{15}N-labelled, monomeric sample of BlgB was obtained from and prepared by Patrick Edwards, New Zealand Dairy Research Institute. The sidechain nuclei of the Trp residues were not isotopically enriched with ^{13}C. The sample was prepared using 50 mM NaOAc buffer of pH 2.6 in 10% D_2O, 90% H_2O (v/v). The protein concentration was approximately 1.5 mM.

3.1.2 C4BP~1,2

A ^{13}C,^{15}N-labelled sample of C4BP~1,2 was obtained from Linda Mark, Lund University, Sweden. Sample preparation was carried out by Huw Jenkins, University of Edinburgh, using 20mM NaOAc buffer (pH 4.5, 0.05% NaN_3) and 10% D_2O, 90% H_2O (v/v) resulting in a protein concentration of ca. 1.5 mM.

3.2 Processing of Spectra

The AZARA program (Boucher, 1993) was used to process spectra. The AZARA processing script contains the various commands used to improve the quality of the final spectrum and the necessary data-handling operations required according to the method

by which the spectrum was acquired. These included zerofilling the data once to the current or next nearest 2^n points, using sinebell squared functions to remove truncation of FIDs and employing trigonometric baseline corrections. Where used, the 0, 0 zero and first order phasing of indirect dimensions made it necessary to scale the first point of the FID in these dimensions by 0.5 to remove the DC offset of the spectrum.

3.2.1 Maximum entropy

Maximum entropy processing (Gull & Daniell, 1978) is an alternative to the Fourier transform method of obtaining a spectrum and is useful for improving the resolution and reducing noise in indirectly detected dimensions of multi-dimensional NMR experiments (Sibisi et al., 1984). This method operates by randomly producing large numbers of spectra and inverse Fourier transforming them to obtain FIDs. The FIDs are compared to the experimental FID to assess the closeness of fit. The FID which best fits the experimental FID and contains the smallest amount of information (and, therefore, the maximum entropy) is then used.

3.3 Resonance Assignment

The resonance assignment of double labelled samples requires assigning all, or as many as possible, of the ^1H, ^{13}C and ^{15}N resonances of the protein. For this purpose numerous triple-resonance 3D experiments have been devised which correlate these nuclei, either via purely INEPT-based experiments or experiments which employ spin-lock sequences. These experiments are commonly based upon a 2D ^{15}N-^1H-HSQC root with the third spectral dimension being used to label ^{13}C or ^1H nuclei. The various stages of resonance assignment and the experiments used in this project are outlined here.

3.3.1 Backbone Assignment

The first stage of resonance assignment involves identifying which of the ^{15}N-^{1}H-HSQC peaks belong to which residue of the protein, essentially establishing the connectivities between the signals of this spectrum. Two 3D experiments were used for this purpose, the CBCA(CO)NH (Grzesiek & Bax, 1992) and HNCACB (Wittekind & Mueller, 1993) experiments. The CBCA(CO)NH experiment correlates the N^H and H^N chemical shifts of residue *i* with the C^β and C^α chemical shifts of residue *i*-1. The HNCACB experiment provides the same correlations but, additionally, includes the C^α and C^β shifts of residue *i*. This difference is due to the differing magnetization transfer pathways in these experiments. In the CBCA(CO)NH, following transfer from H^{Ni} to N^{Hi}, magnetization is transferred from N^{Hi} to $C^{\alpha i-1}$ by stepping through C'^{i-1} via the $^{1}J_{NC'}$ and $^{1}J_{C'C\alpha}$ couplings. In the HNCACB experiment, stepping through C' does not occur and instead, following initial transfer from H^{Ni} to N^{Hi}, the magnetization is transferred via the $^{1}J_{C'C\alpha}$ and $^{2}J_{C'C\alpha}$ couplings to *both* $C^{\alpha i}$ and $C^{\alpha i-1}$ respectively. In both cases, the magnetization is subsequently transferred to C^β. Thus, a strip of the CBCA(CO)NH spectrum (at a particular ^{15}N and ^{1}H position) will contain two peaks and the equivalent strip of the HNCACB spectrum will contain four peaks as illustrated in Figure 3.1.

The CBCA(CO)NH can therefore be used to indicate which of the four HNCACB peaks are from residue *i* and which from residue *i*-1. The connectivities between the ^{15}N-^{1}H-HSQC peaks can then be established by matching $C^{\alpha i}$ and $C^{\beta i}$ shifts of a strip with $C^{\alpha i-1}$ and $C^{\beta i-1}$ shifts of another strip. This principle is illustrated in Figure 3.2.

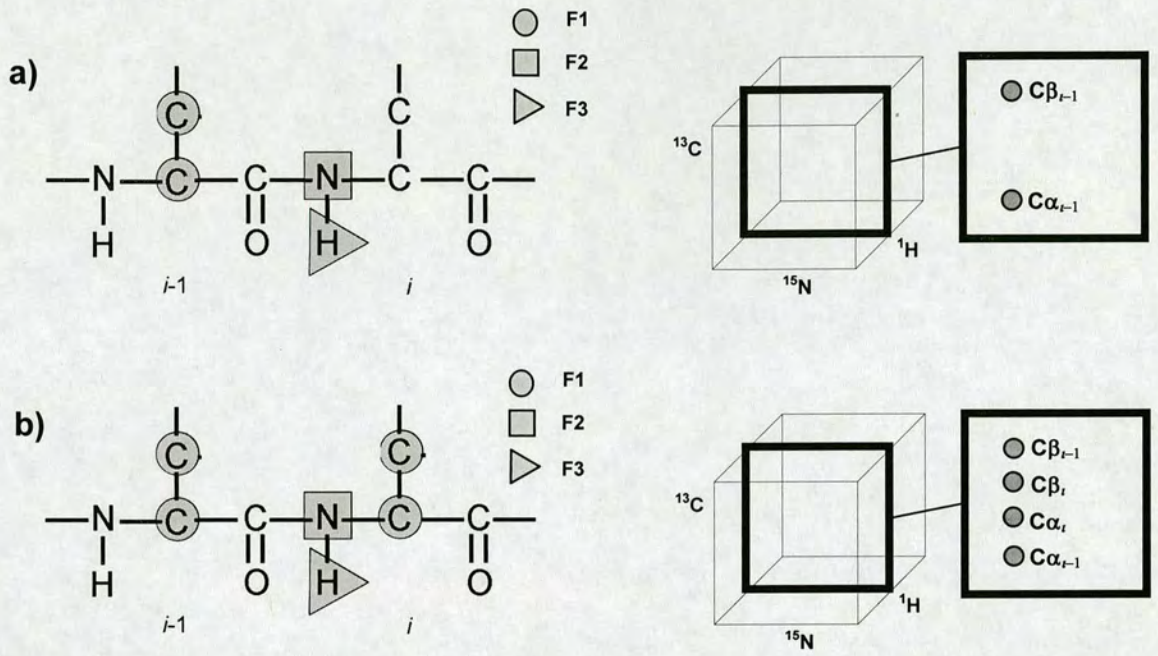


Figure 3.1 Representations of a) the CBCACONH experiment and b) the HNCACB experiment. In each case, the nuclei labeled in each dimension are shown along with a cartoon of a particular strip.

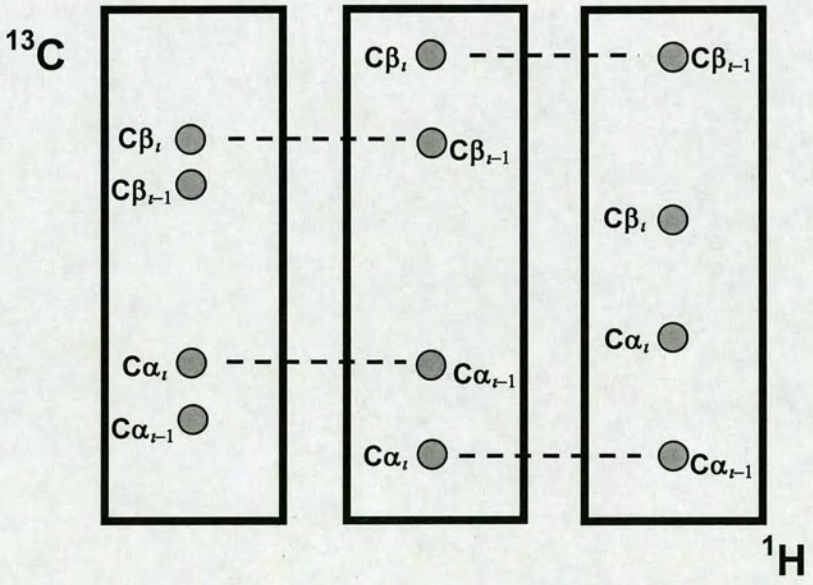


Figure 3.2 Cartoon representations of strips of the HNCACB spectrum. Dashed lines indicate connectivities established by chemical shift matches.

Once connectivities have been established, it still remains to map these connections onto the amino acid sequence. This can be done using known values of C^{α} and C^{β} chemical shifts for particular residues obtained from a chemical shift database to identify the residue type or types which may have given rise to a particular strip. The residue type and connectivity information can then be matched with the amino acid sequence.

Obviously, proline residues will not give rise to any peaks in such spectra as they do not contain the amide protons through which the signals are detected. This leads to 'breaks' in the sequential assignment at the proline positions. Such breaks can also be of use in mapping assignments onto the amino acid sequence.

A further experiment, HBHA(CO)NH (Grzesiek & Bax, 1993b) is used to assign H^{β} and H^{α} chemical shifts. This experiment is analogous with the CBCA(CO)NH experiment but with ^{1}H frequencies (of $H^{\alpha i-1}$ and $H^{\beta i-1}$) in F_1 as shown in Figure 3.3.

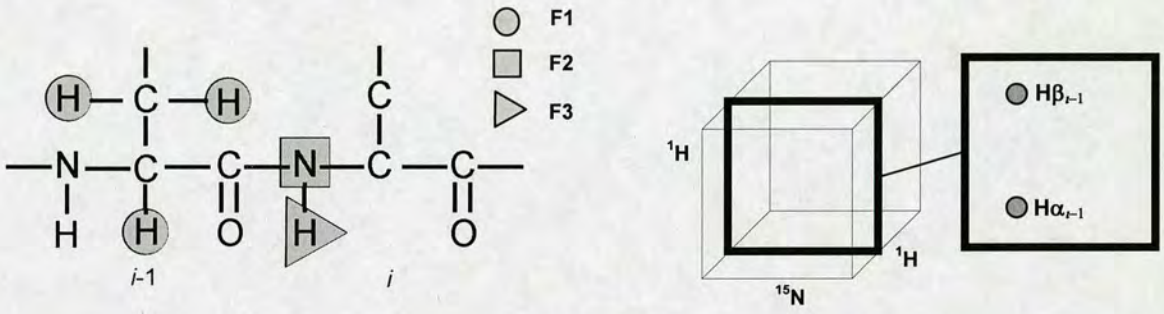


Figure 3.3 Representations of the HBHA(CO)NH experiment

If required, a HBHANH experiment, analogous to the HNCACB experiment, can be acquired which also contains $H^{\alpha i}$ and $H^{\beta i}$ resonances thus aiding sequential assignment. In this project, however, the ^{13}C spectra were sufficient for this purpose.

3.3.2 Sidechain Assignment

For the assignment of the sidechain resonances, two 3D experiments were employed – the H(C)(CO)NH (Logan et al., 1993) and (H)C(CO)NH (Grzesiek et al., 1993) experiments. These experiments correlate either all *i*-1 ^{1}H or all *i*-1 ^{13}C resonances with the N^{H} and H^{N} resonances of residue *i* as shown in Figure 3.4.

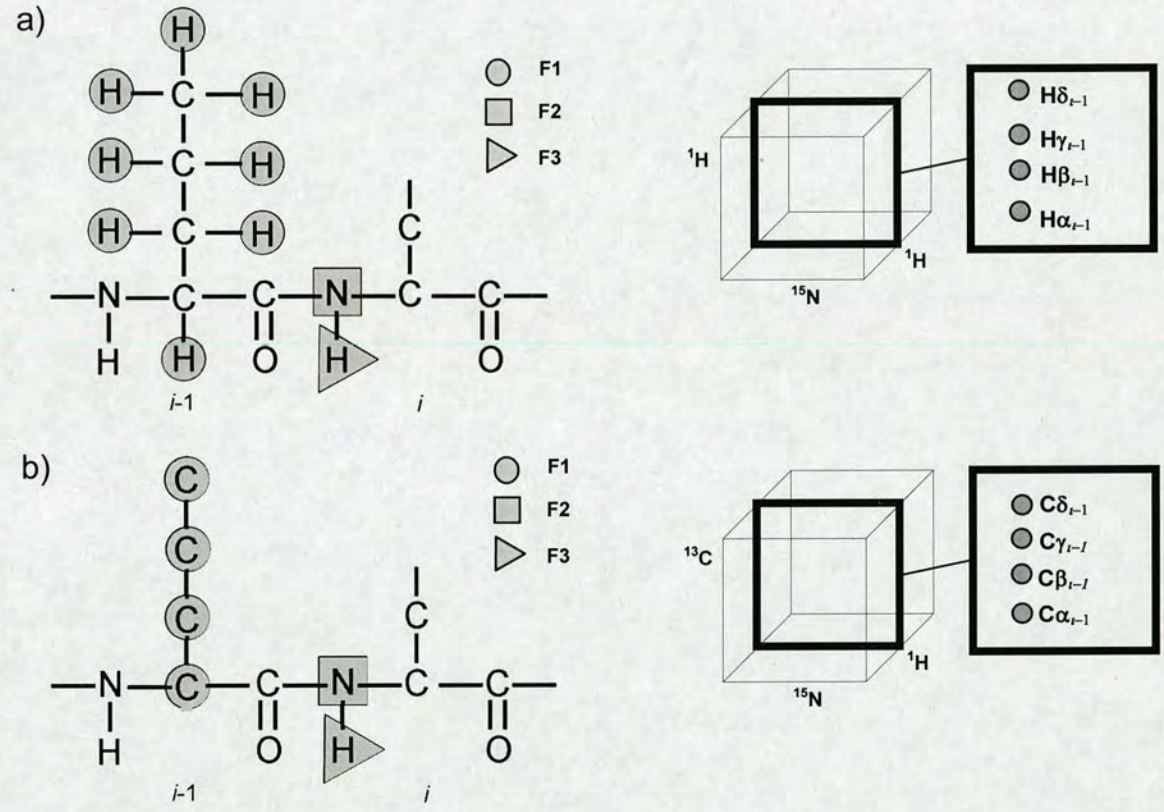


Figure 3.4 Representations of a) the H(C)(CO)NH and b) the (H)C(CO)NH experiments.

These experiments can also be of use in identifying residue types during sequential assignment by identifying peak patterns indicative of particular residues and, again, by referring to a database of known chemical shifts. As was the case in the previous experiments, proline residues are absent and, therefore, the sidechain resonances of pre-proline residues cannot be assigned. It may not be possible to assign some resonances

due to signal overlap. These problems can be addressed by acquiring a third, 3D sidechain experiment – the HCCH-TOCSY (Kay et al., 1993). This experiment provides ^{1}H spin system information similar to that of a 2D ^{1}H-^{1}H TOCSY experiment but, by using a ^{13}C spinlock instead of ^{1}H and labeling the ^{13}C, the spectrum is edited along a third axis according to the frequency of the ^{13}C nuclei. Assignment of this spectrum also serves to provide somewhat more accurate chemical shifts for those resonances already assigned.

NOESY spectra are also used in the assignment process (Wuethrich, 1986). The most intense crosspeaks in a strip will usually be due to intra-residue NOEs or NOEs between adjacent residues. This information can be used to confirm the assignments obtained using the above experiments, particularly in the case of pre-proline residues. The amide ^{1}H's of Gln and Asn are assigned, following assignment of the other resonances, by identifying the intense NOEs to either H^{γ} (Gln) or H^{β} (Asn). Assignment of amine groups can also achieved in this way but these signals are seldom observed due to the amine ^{1}H's being highly labile.

3.3.3 Aromatic Assignment

Assignment of the aromatic ^{1}H and ^{13}C resonances is achieved using a series of 2D experiments. The INEPT transfer steps of the above resonance assignment experiments are optimised for $^{1}J_{CH}$ and $^{1}J_{CC}$ values of aliphatic nuclei which are significantly different from those of aromatics. Also, the ^{13}C spin-lock of e.g. HCCH-TOCSY experiment does not cover aromatic resonances. Thus, aromatic-specific experiments are required. The first two of these experiments correlate C^{β} resonances of aromatic residues with either the H^{δ} or H^{ε} resonances of the aromatic ring (Yamazaki et al., 1993) as shown in Figure 3.5 a) and b). A third experiment, shown in Figure 3.5 c), involves a direct correlation of aromatic protons with the ^{13}C nuclei to which they are bound.

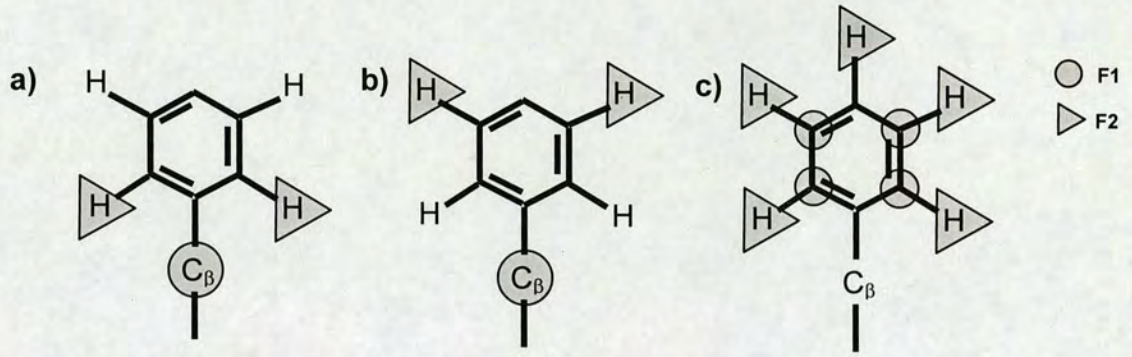


Figure 3.5 Schematic representations of a) the (HB)CB(CGCD)HD, b) the (HB)CB(CGCDCE)HE and c) the aromatic HSQC experiments

3.4 Structure Calculation

Several programs have been developed for the purpose of calculating structures from NMR-derived distance restraints using various algorithms and detailed reviews of the process exist elsewhere (Nilges, 1996, 1998; Guentert, 2003).

In this project, structure calculations were performed with the CNS program (Brunger et al., 1998), incorporating the ARIA protocol (Nilges et al., 1997) for assignment of ambiguous NOE's. The CNS structure calculation protocol operates using an energy minimisation procedure where the total energy of the system, E_{tot}, is the sum of various individual energy terms that define the system.

$$E_{tot} = \sum(E_{bond} + E_{angle} + E_{dihe} + E_{impr} + E_{vdw} + E_{NOE})$$

Where *bond* refers to covalent bonds; *angle* refers to bond angles; *dihe* refers to dihedral angles; *impr* refers to improper angles; *vdw* refers to van der Waals interactions and *NOE* refers to the contribution from the NOE restraints.

For structure determination by NMR the most significant input is the distance restraints obtained from NOESY experiments. Intensities of crosspeaks which have been picked in the NOESY spectra were automatically calculated within the Analysis program. These intensities were then converted into distance restraints by grouping crosspeaks together into intensity bands corresponding to particular distance bounds. The intensity bands and distance classes used are given in Table 3.1.

Intensity (I)	**Distance (Å)**
I>3.0	0 - 2.7
3.0>I>1.5	0 - 3.3
1.5>I>0.3	0 - 5
0.3>I>0.1	0 - 6

Table 3.1 NOE distance classes

Only upper bounds were applied to the distance classes since the intensity of a crosspeak is affected by various other factors such as relaxation and spin diffusion and, therefore, a crosspeak may appear weaker than would be expected, were internuclear distances the only consideration.

Once these distance restraint tables have been generated, they are used as input in the CNS structure calculation. Along with the experimental restraint input, a set of empirically derived data including bond angles, stereochemistry, steric interactions and electrostatics known as the 'forcefield' is also used to define the behaviour of the atoms. The PARALLHDG forcefield (Linge & Nilges, 1999) was used in this project.

The calculation itself essentially consists of simulated annealing of the protein – heating the system to 2000K and then slowly cooling it to 100K. Three stages of annealing are performed with the results of the previous stage used as an input for the subsequent stages. In the final annealing step, prochiral methylene atoms with non-stereospecific

assignments have their assignments swapped and the lower energy version of each prochiral centre is selected. At the end of the calculation process a user-defined number of structures are produced. These structures are then used, via the ARIA protocol, to refine the input restraints prior to another round of structure calculation.

3.4.1 Violations

For each structure which is calculated a record is made of the experimental restraints which have been violated i.e. those which have not been satisfied in the structure. Violations may occur for a number of reasons – the peak may be incorrectly assigned, there may be several contributions to a peak, the peak may in fact be some artifact or noise or the structure may, in fact, be incorrect. For NOE restraints a restraint was considered to be violated if it was >0.3Å beyond the bound defined in the restraint tables. Violations identified in this way are checked by returning to the spectra and re-examining the peak which gave rise to the restraint to correct its assignment, make it more ambiguous or, in the case of artefacts, remove the peak altogether.

3.4.2 ARIA

As described previously, the ARIA protocol allows refinement of ambiguous distance restraints via comparison with the results of the CNS structure calculation. The structures which are to be used in this process are chosen on the basis of lowest total energy. The choice of structures is significant as it determines the effectiveness of the calculation procedure. In any calculation performed with sufficient 'correct' restraint information, an ensemble of structures will be generated which share the same energy and overall fold. Examination of a plot of the E_{total} and E_{NOE} of an ensemble will show a plateau corresponding to these structures and these are selected for use in the ARIA protocol. In addition to these criteria, the backbone rmsds of a progressively larger subset or 'cluster' of the total structure ensemble (ranked according to E_{total}) is produced (starting with the two lowest energy structures and successively adding the next highest

energy structure) to show the variation of rmsd with the increasing ensemble size, thus providing a further criterion by which to judge the cut-off point (Diamond, 1995).

Filtering

Using the ensemble of structures, ambiguous restraints are filtered. The contribution of each ambiguous restraint is defined according to the proportion of the intensity of the peak which could be attributed to that restraint (according to the distances derived from the structure ensemble). Restraints which contribute less than 1% of the intensity of a peak are discarded.

Checking

After filtering of the restraints, the new restraint tables (both unambiguous and ambiguous) are submitted to the checking procedure which examined the tables to identify instances of duplication. Duplicate restraints will occur where both symmetry-related peaks of an NOE interaction have been picked and assigned. Where duplicate restraints are found, the restraint with the higher intensity was retained. Once these stages are complete, a new table of distance restraints is generated and a new structure calculation round is initiated using these modified restraints. Eventually, the quality of structures produced will show no improvement over the previous round and it can be assumed that the calculation is complete.

3.4.3 Water refinement

The final stage of the structure calculation was to refine the converged structures by recalculating the structures in water. This was performed using the RECOORD database (Nederveen et al., 2005) implemented with CNS. Recalculating structures in explicit solvent has been shown to give significant improvements to measures of structural quality or soundness such as Ramachandran plot statistics (Spronk et al., 2004).

3.4.4 Quality Checks

The quality of experimental data and validation of the resulting structures is an important consideration in solving protein structures by NMR.

Recently, the program QUEEN was developed (Nabuurs et al., 2003) which assesses the uniqueness of experimental distance restraints used in a structure calculation by evaluating the extent to which the addition of an individual restraint reduces the total structural uncertainty. This results in an output wherein the uniqueness of a particular piece of information (I_{uni}) is expressed as a fraction of the total information for the system (I_{total}). In this way, the most unique information present in a set of restraint data can be identified i.e. those restraints which are least supported by other restraints (experimental or empirical). Restraints identified as being particularly unique are more likely to have a significant effect upon the structure calculation and should therefore be examined to assess their validity.

The quality of structures solved by NMR can be assessed by comparing various structural parameters against databases of values from high resolution X-ray structures (Spronk et al., 2004; Nabuurs et al, 2004, 2005). Bond lengths, bond angles, torsion angles, atomic contacts and various other parameters can be used to assess the 'normality' of the protein structure and, therefore, give a measure of how 'good' the structure is. Various programs and databases exist for this purpose. For this project, the programs PROCHECK (Laskowski et al., 1993) and WHATIF (Vriend, 1990) were employed to assess the quality of structures.

The program PROCHECK was use to obtain Ramachandran (Ramachandran & Sasiskharan, 1968) plots which show the distribution of ϑ/ψ angles within a protein. Certain combinations of these angles are more favoured than others and, thus, the plot itself can be divided into regions of favourable and unfavourable angles. Good structures will have the majority of their ϑ/ψ angles in the most favoured regions.

Chapter 4

Complementary Spectra

4.1 Introduction

Four complementary experiments were chosen to be investigated as potential assignment aids. These were the 3D $^{13}C,^{15}N$-HSQC-NOESY-HSQC (CNH-NOESY) and $^{15}N,^{15}N$-HSQC-NOESY-HSQC (NNH-NOESY) for use with the 3D ^{15}N-edited-NOESY (HNH-NOESY); the $^{13}C,^{13}C$-HSQC-NOESY-HSQC (CCH-NOESY) for use with the ^{13}C-edited-NOESY (HCH-NOESY); the $^{13}C,^{13}CH_3$-HSQC-NOESY-HSQC (CCH_3-NOESY) for use with the $^{13}CH_3$-edited-NOESY (HCH_3-NOESY).

It was decided not to acquire the NOESY experiments in a time-shared manner (Section 2.3.1:p32) to ensure that all experiments benefited from the maximum sensitivity and resolution. However, if desired, the application of time-shared labeling periods can be achieved with simple extensions of the experiments described.

Pulse programs for these experiments were written and implemented on a Bruker 600 MHz NMR spectrometer fitted with a 5mm triple-axis gradient, triple resonance probe. There follows a brief description of each of the complementary experiments.

4.2 CNH-NOESY - Figure 4.1

As with all the complementary experiments, this sequence begins with an HSQC. An initial ^{13}C $\pi/2$ pulse, used to destroy any magnetization which originates on the ^{13}C nuclei, is followed by ^{13}C HSQC. PFGs are applied on either side of the ^{13}C and ^{1}H INEPT π pulses to remove any magnetization which did not experience a complete 180° rotation and a gradient pulse is applied while the H_ZC_Z coherence exists prior to the polarization transfer pulse (Muhandiram et al., 1993) generating the H_zC_y coherence. Due to difficulties with spectral folding in this dimension, if aromatic resonances were also to be labelled, it would be necessary to have much longer acquisition times. In many cases, the number of aromatic proton signals will be far fewer than those of the aliphatic protons and sufficiently resolved to allow unambiguous assignment of their signals. We have therefore opted for using ^{13}C π pulses that invert only aliphatic carbons thus discarding the signal from aromatic carbons. For this purpose, a Q5 pulse (Emsley & Bodenhausen, 1990) was used of duration 160μs. The inversion profile of the pulse is given in Figure 4.2. If aromatic signals were of interest, the INEPT π pulses could be replaced with adiabatic inversion pulses, to optimize the transfer of magnetization via differing couplings ($^{1}JCH_{ali}$ and $^{1}JCH_{aro}$).

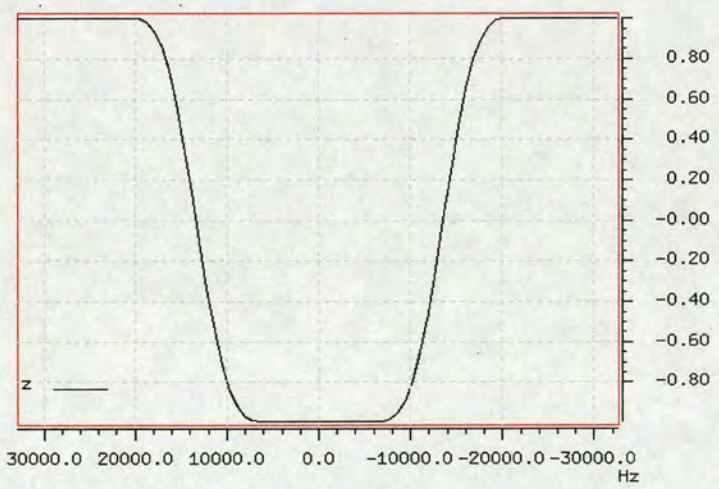


Figure 4.2: Inversion profile of the ^{13}C π pulse for selective excitation of aliphatic carbons

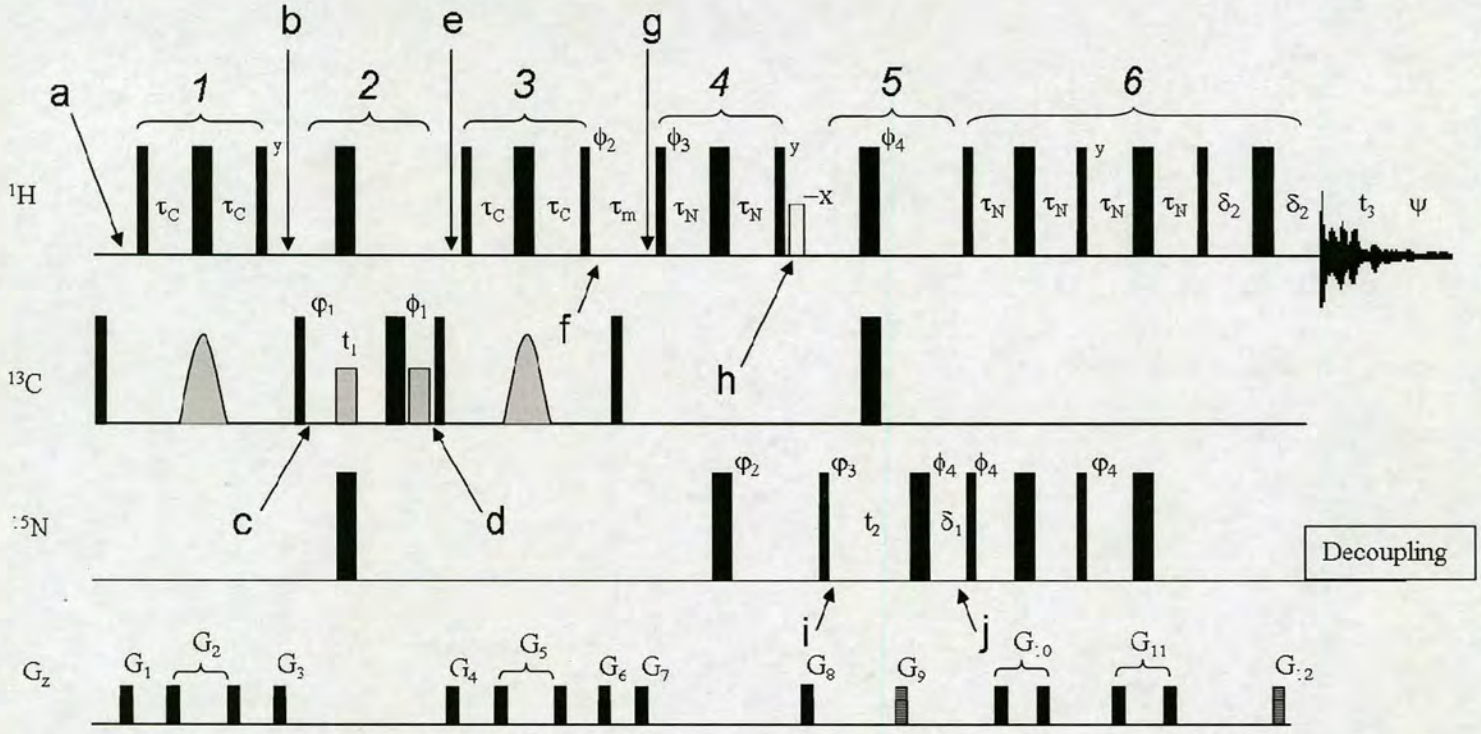


Figure 4.1: The 3D ^{13}C,^{15}N-HSQC-NOESY-HSQC sequence. φ_1=4x, 4-x; φ_2=x; φ_3= x, -x; φ_4=2y, 2-y; ϕ_1=4y,4-y; ϕ_2=8y,8-y; ϕ_3=8x,8-x; ϕ_4=2x,2-x; ψ=x,-x,-x,x,-x,x,x,-x,-x,x,-x,x,x,-x,x,-x,x. φ4 shifted by 180° for imaginary component. All other pulses are applied from x axis. $\delta_1=\tau_g$ | $\tau_{C\pi}$, $\delta_2=\tau_g$, where τ_g=length of the PFG and $t_{C\pi}$=length of the ^{13}C π pulse. $\tau_C=1/(4x^1J_{CH})$, $t_N=1/(4x^1J_{NH})$, τ_m=mixing time. Grey bars are 190μs Q5 pulses for selective ^{13}CO inversion; grey shaped pulses are selective aliphatic inversion pulses (see text). The white bar is a 1400μs rectangular pulse for selective water excitation. All PFG's are 1ms and applied with the following strengths (%): G_1=7, G_2=13, G_3=23, G_4=-18, G_5=27, G_6=30, G_7=17, G_8=-50, G_9=80, G_{10}=11, G_{12}=8.1 The grey G_z pulses are applied in the ratio 10:1.

Product Operator Description:

$$\mathbf{a}\ H_{1z} \xrightarrow{\mathbf{1}} \mathbf{b}\ 2H_{1z}C_z \xrightarrow{90°_C} \mathbf{c}\ -2H_{1z}C_y \xrightarrow{\mathbf{2}} \mathbf{d}\ -2H_{1z}C_y c_{\Omega C t1} \xrightarrow{90°_C} \mathbf{e}\ -2H_{1z}C_z c_{\Omega t1} \xrightarrow{\mathbf{3}} \mathbf{f}\ H_{1z} c_{\Omega C t1}$$

$$\xrightarrow{\tau_m} \mathbf{g}\ H_{2z} c_{\Omega C t1} \xrightarrow{\mathbf{4}} \mathbf{h}\ 2H_{2z}N_z c_{\Omega C t1} \xrightarrow{90°_N} \mathbf{i}\ -2H_{2z}N_y c_{\Omega C t1} \xrightarrow{\mathbf{5}} \mathbf{j}\ -2H_{2z}N_y c_{\Omega C t1} c_{\Omega N t2} + 2H_{2z}N_x c_{\Omega C t1} s_{\Omega N t2}$$

$$\xrightarrow{\mathbf{6}} \mathbf{k}\ \begin{cases} -H_{2y} c_{\Omega C t1} c_{\Omega N t2} + H_{2x} c_{\Omega C t1} s_{\Omega N t2} & \text{(real)} \\ -H_{2y} c_{\Omega C t1} c_{\Omega N t2} - H_{2x} c_{\Omega C t1} s_{\Omega N t2} & \text{(imaginary)} \end{cases}$$

Where $c_{\Omega C t1}=\cos\Omega_C t_1$, $c_{\Omega N t2}=\cos\Omega_N t_2$ and $s_{\Omega N t2}=\sin\Omega_N t_2$

^{13}C chemical shift labeling is then achieved using the variable time method. ^{1}H, ^{15}N and $^{13}C'$ π pulses are applied in the centre of the t_1 delay to refocus the couplings to these nuclei. A short ^{13}C spin echo is incorporated by applying a ^{13}C π pulse followed by another $^{13}C'$ π pulse to remove the effects of any ^{13}C frequency shift which occurred due to the first $^{13}C'$ pulse (Bloch-Siegert shift). Again, a gradient pulse is applied before returning the magnetization to ^{1}H. At the end of the refocused INEPT, a further ^{1}H $\pi/2$ pulse returns the magnetization to the z-axis to allow transfer of magnetization via NOE during the mixing period, τ_m. A ^{13}C $\pi/2$ pulse during τ_m removes any magnetization that was not fully refocused by converting any remaining $2H_xC_z$ coherence into the unobservable, multiple quantum coherence $2H_xC_y$. Since the desired magnetization resides on the z-axis at this point, a PFG is also applied during τ_m to remove any other, unwanted magnetization. Following the NOE transfer, the second, ^{15}N HSQC commences. As with the ^{13}C HSQC, zz gradients are applied prior to the magnetization transfer pulses and decoupling of the ^{1}H and ^{13}C nuclei is achieved by applying π pulses to these nuclei during t_2. A short spin echo is employed following t_2 to account for any additional chemical shift evolution which occurs during the application of the encoding gradient and the relatively long ^{13}C π pulse. To improve the sensitivity of this sequence, the final, refocused INEPT step is replaced with a sensitivity enhancement block (Cavanagh et al., 1991). Gradient selection is used during this block to remove the water signal. In addition, a selective 90°, water flip-back pulse is applied prior to the t_2 period to ensure that the water signal resides on the z-axis prior to acquisition. The initial values of t_1 and t_2 are set to half the value of the time increments in these dimensions to give zero and first order phasing of 90 and -180 in these dimensions. This allows pulses applied to other nuclei to be accommodated by t_1 and gives smoother baselines in the spectrum.

Since the magnetization passes first through ^{13}C nuclei and then ^{15}N nuclei before detection through the ^{15}NH nuclei, the magnetization which originated from the ^{15}NH nuclei is not detected and, thus, the spectrum contains no diagonal peaks. As can be seen

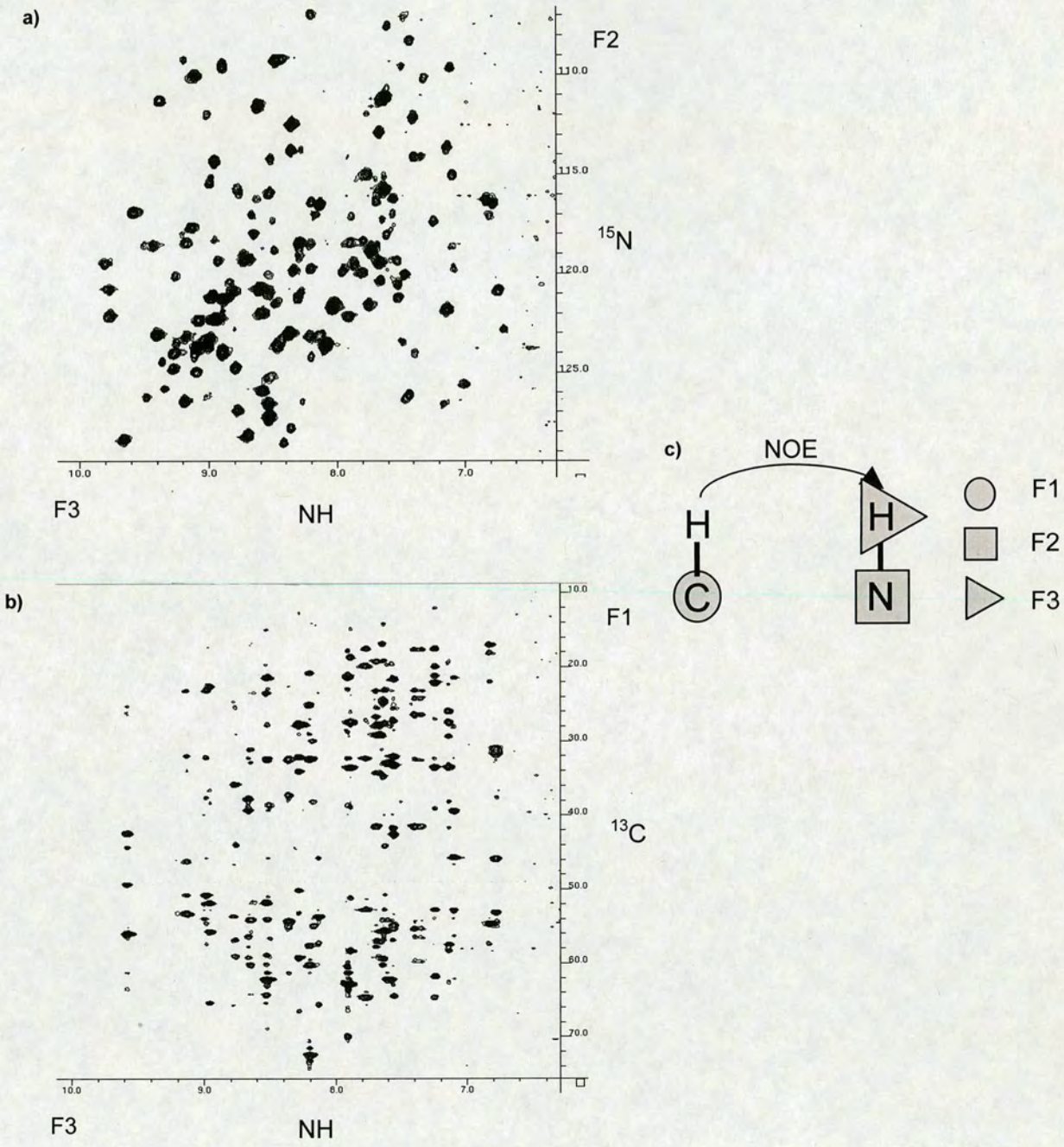


Figure 4.3: a) F_2F_3 and, b), F_1F_3 projections of the CNH-NOESY spectrum of BlgB. c) A schematic illustration of the experiment

from the F_1F_3 and F_2F_3 projections in figures 4.3 a) and b), the spectrum therefore superficially resembles the (H)C(CO)NH spectrum, but with ^{13}C resonances not limited to *i*-1 residues.

4.3 NNH-NOESY - Figure 4.4

This sequence is very similar to the CNH-NOESY with the difference that the first HSQC step transfers magnetization to the ^{15}N nuclei. Chemical shift labeling is achieved via the variable time method in both indirect dimensions with short spin-echos used in t_1 and t_2 to compensate for chemical shift evolution occurring during the decoupling ^{13}C pulse. Water suppression is achieved by the use of selective water pulses to return the water signal to the z axis prior to acquisition. A WATERGATE sequence is also employed in the final, refocusing INEPT step to further reduce the detected water signal. Initial incrementation of t_1 and t_2 by half of the increment in these dimensions is also used.

In this case, since the magnetization both originates on and is detected through $^{15}N^H$ nuclei, the magnetization which is not transferred during τ_m gives rise to intense autocorrelation peaks in the spectrum as shown in Figure 4.5. No benefit was obtained by using a sensitivity enhancement block in this experiment and so the shorter, simpler sequence was employed.

4.4 CCH-NOESY – Figure 4.6

In this complementary experiment two ^{13}C HSQCs are separated by the NOESY mixing time. PFGs are applied at points where the H_zC_z coherences are present and further PFGs sandwich the π pulses of the INEPT steps. Evolution of the coupling to ^1H, ^{15}N and $^{13}C'$ nuclei during t_1 and t_2 is removed by applying π pulses on these nuclei, as in the previous experiments, necessitating the use of a spin echo and additional $^{13}C'$ pulse

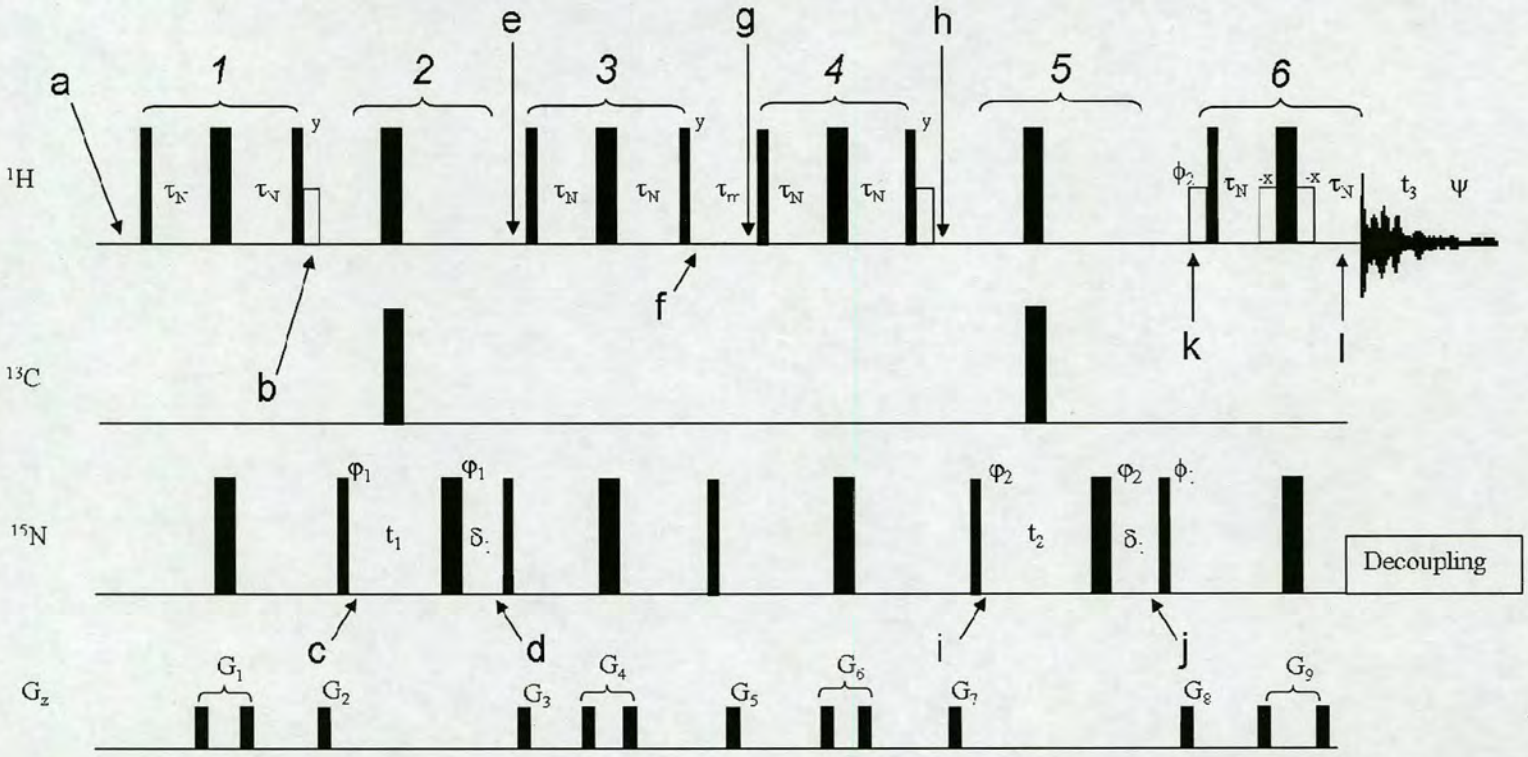


Figure 4.4: The 3D ^{15}N,^{15}N-HSQC-NOESY-HSQC sequence. $\pi/2$ and π pulses are represented by thin and thick black bars respectively. ϕ_1=8x, 8-x; ϕ_2=2x,2-x; φ_1= x,-x; φ_2= 4x, 4-x; ψ= x, -x, -x, x, -x, x, x, -x, -x, x, x, -x, x, -x, -x, x. All other pulses are applied from x axis. $\delta_1 = \tau_{C\pi}$, where $\tau_{C\pi}$ is the length of the ^{13}C π pulse. $\tau_N = 1/(4x^1J_{N-H})$, τ_m=mixing time. All PFGs are 1ms long and applied with the following strengths (%): G_1=14, G_2=30, G_3=18, G_4=27, G_5=40, G_6=5, G_7=12, G_8=60, G_9=-60. Sign discrimination in F_1 and F_3 is achieved via the States-TPPI method by incrementing the phases of φ_1 and φ_2 by 90°. White bars are 1400μs rectangular pulses for selective water excitation.

Product Operator Description:

$$\underset{a}{H_{1z}} \xrightarrow{1} \underset{b}{2H_{1z}N_z} \xrightarrow{90°_N} \underset{c}{-2H_{1z}N_y} \xrightarrow{2} \underset{d}{-2H_{1z}N_y c_{\Omega Nt1}} \xrightarrow{90°_N} \underset{e}{-2H_{1z}N_z c_{\Omega Nt1}} \xrightarrow{3} \underset{f}{H_{1z}c_{\Omega Nt1}}$$

$$\xrightarrow{\tau_m} \underset{g}{H_{2z}c_{\Omega Nt1}} \xrightarrow{4} \underset{h}{2H_{2z}N_z c_{\Omega Nt1}} \xrightarrow{90°_N} \underset{i}{-2H_{2z}N_y c_{\Omega Nt1}} \xrightarrow{5} \underset{j}{-2H_{2z}N_y c_{\Omega Nt1} c_{\Omega Nt2}} \xrightarrow{90°_N} \underset{k}{-2H_{2z}N_z c_{\Omega Nt1} c_{\Omega Nt2}} \xrightarrow{6} \underset{l}{-H_{2x}c_{\Omega Nt1} c_{\Omega Nt2}}$$

Where $c_{\Omega Nt1} = \cos\Omega N t_1$ and $c_{\Omega Nt2} = \cos\Omega N t_1$

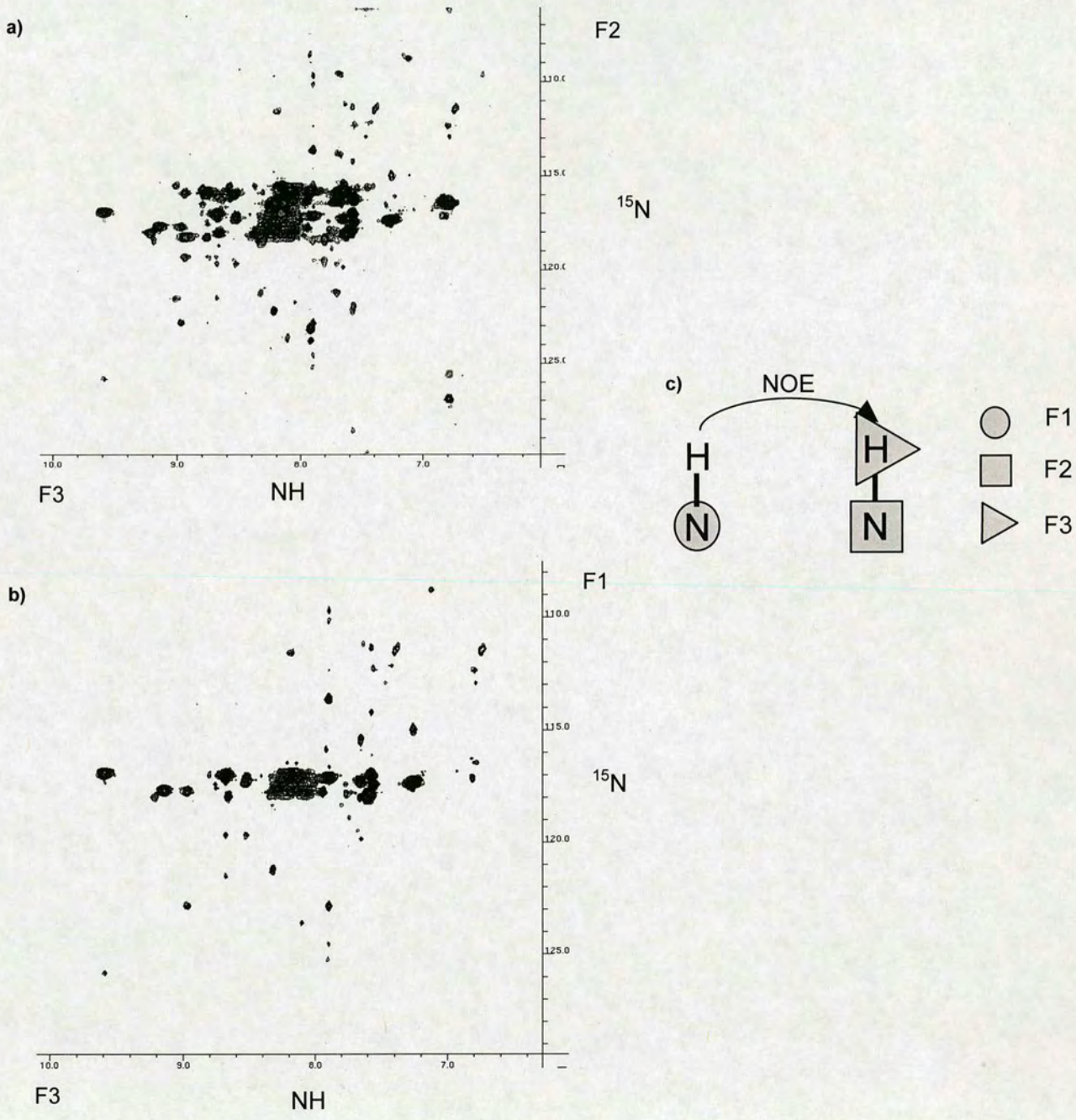


Figure 4.5: a) F_2F_3 and, b), F_1F_3 projections of the NNH-NOESY spectrum of BlgB. c) A schematic illustration of the experiment

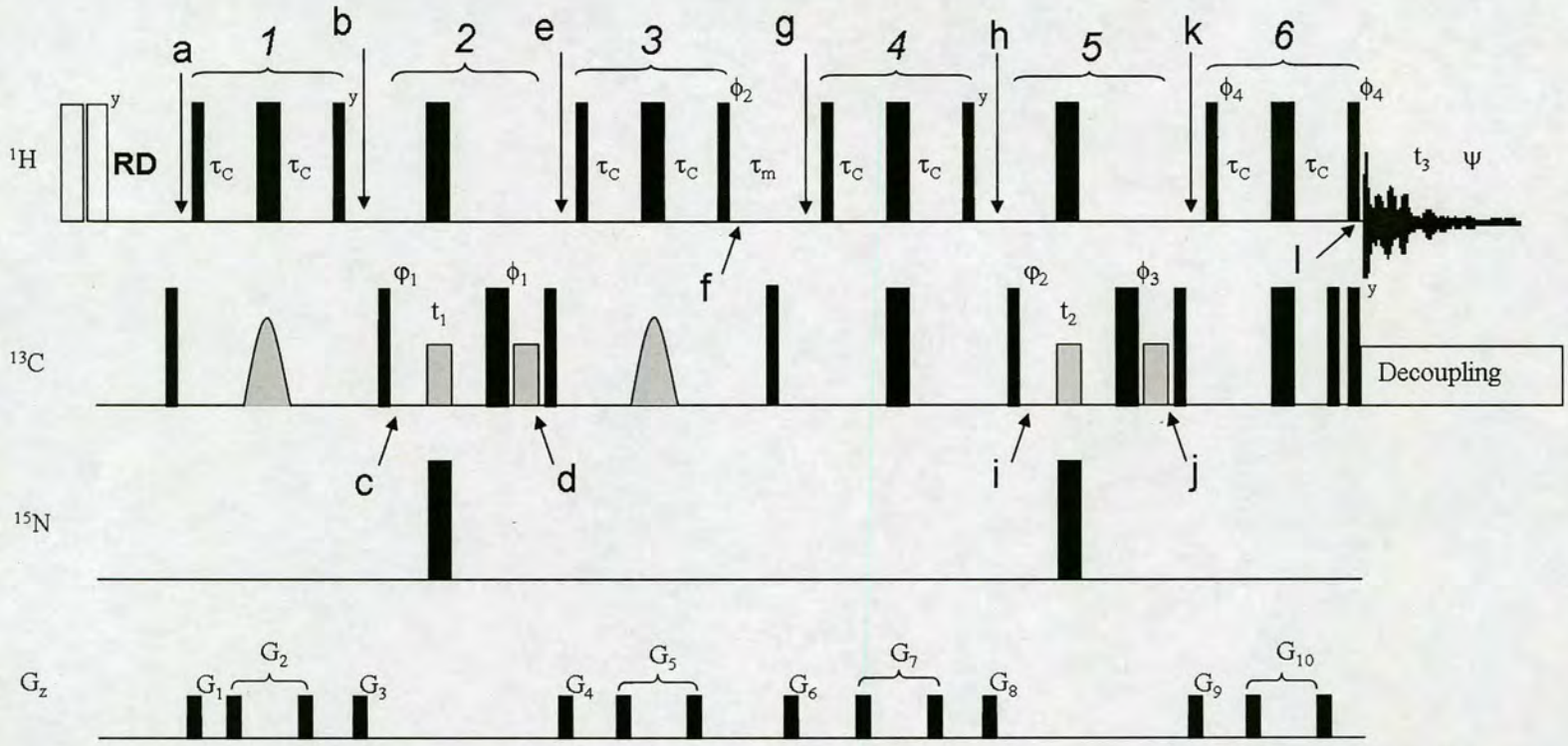


Figure 4.6: The 3D ^{13}C,^{13}C-HSQC-NOESY-HSQC sequence. φ_1=x, -x; φ_2=2x, 2-x ϕ_1=4y,4-y; ϕ_2=16y,16-y; ϕ_3=4y, 4-y; ϕ_4=4x,4y,4-x,4-y; ψ=x,-x,-x,x,-y,y,y,-y,-x,x,x,-x,y,-y,-y,y,x,x,-x,y,-y,-y,y,x,-x,-x,x,-y,y,y,-y. All other pulses are x. τ_C=1/(4x$^1J_{CH}$), τ_m=mixing time, RD=relaxation delay of 1.5s. PFG's are 1ms long and applied with the following strengths (%): G_1=30, G_2=10, G_3=13, G_4=-15, G_5=5, G_6=27, G_7=33, G_8=80, G_9=-22, G_{10}=12. Grey bars are 190μs Q5 pulses for selective ^{13}CO inversion; grey shaped pulses are selective aliphatic carbon inversion pulses (see text), the white bars are rectangular pulses of length 2.5 ms for removal of coherent magnetization. Sign discrimination in F_1 and F_3 is achieved via the States-TPPI method by incrementing the phases of φ_1 and φ_2 by 90°.

Product Operator Description:

$$\mathbf{a}\ \mathbf{1}\ \mathbf{b}\ {}^{90°_C}\ \mathbf{c}\ \mathbf{2}\ \mathbf{d}\ {}^{90°_C}\ \mathbf{e}\ \mathbf{3}\ \mathbf{f}$$
$$H_{1z} \rightarrow 2H_{1z}C_z \rightarrow -2H_{1z}C_y \rightarrow -2H_{1z}C_y c_{\Omega Ct1} \rightarrow -2H_{1z}C_z c_{\Omega Ct1} \rightarrow H_{1z} c_{\Omega Ct1}$$
$$\mathbf{g}\ \mathbf{4}\ \mathbf{h}\ {}^{90°_C}\ \mathbf{i}\ \mathbf{5}\ \mathbf{j}\ {}^{90°_C}\ \mathbf{k}\ \mathbf{6}\ \mathbf{l}$$
$$\xrightarrow{\tau_m} H_{2z}c_{\Omega Ct1} \rightarrow 2H_{2z}C_z c_{\Omega Ct1} \rightarrow -2H_{2z}C_y c_{\Omega Ct1} \rightarrow -2H_{2z}C_y c_{\Omega Ct1} c_{\Omega Ct2} \rightarrow -2H_{2z}C_z c_{\Omega Ct1} c_{\Omega Ct2} \rightarrow -H_{2x} c_{\Omega Ct1} c_{\Omega Ct2}$$

Where $c_{\Omega Ct1}$=cos$\Omega_C t_1$ and $c_{\Omega Ct2}$=cos$\Omega_C t_2$

following the t_1 and t_2 periods to eliminate Bloch-Siegert shift. In the de- and refocusing steps of the first HSQC, magnetization is transferred only to and from aliphatic carbons by the use of a selective ^{13}C π pulse. In the final HSQC, magnetization is transferred to both aliphatic and aromatic carbons. Adiabatic ^{13}C π pulses are therefore used to maximize the magnetization transfer through these differing couplings. Treatment of the water signal is effected through phase cycling and a hard water flip-back pulse applied prior to acquisition. A train of two ^{13}C $\pi/2$ pulses is applied immediately before acquisition, acting as a composite $\pi/2$ pulse to convert any remaining $2H_yS_z$ coherence into multiple quantum coherence. Water suppression is achieved by the combined use of phase cycling, gradients prior to polarization transfer pulses and the final $\pi/2$ ^1H pulse which returns the water signal to the z axis (Grzesiek & Bax, 1993a).

As with the NNH-NOESY spectrum, autocorrelation peaks are present, visible as the very intense peaks in the F_1F_3 planes of the spectrum as shown in Figure 4.7. Folding of the F_2 dimension is employed but the spectrum is unfolded in the F_1 dimension.

4.5 CCH_3-NOESY - Figure 4.8

Since methyl-containing sidechains are abundant in proteins, particularly in hydrophobic cores or cavities, methyl-selective experiments have been shown to be useful in protein NMR in both resonance assignment (Uhrin et al., 2000a) and NOESY experiments (Zwahlen et al., 1998; Uhrin et al, 2000b, Xia et al., 2001). By selecting only a small region of the ^{13}C frequencies, the spectral width and, thus, resolution of these experiments can be much greater. Furthermore, the more favourable relaxation properties of methyl carbons allows the use of constant-time labeling, further improving the spectral resolution.

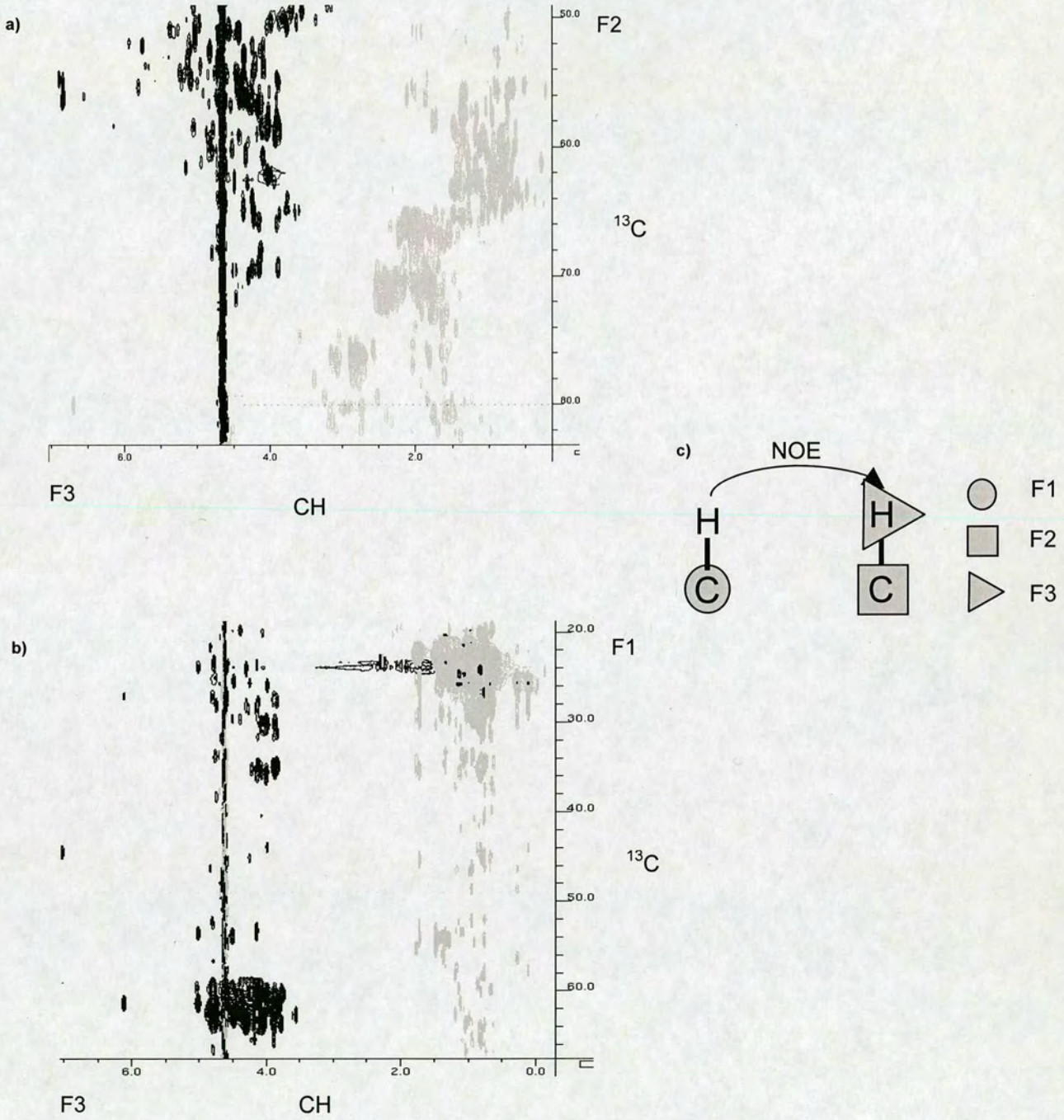


Figure 4.7: a) F_2F_3 and, b), F_1F_3 projections of the CCH-NOESY spectrum of BlgB. c) A schematic illustration of the experiment

This experiment proceeds in a manner identical to the 3D HCH_3-NOESY experiment (Uhrin et al, 2000b) but with an initial ^{13}C HSQC during which the magnetization is labeled with the ^{13}C frequencies in a variable time manner. The selective ^{13}C π pulse used in the INEPT steps was again chosen to cover only aliphatic ^{13}C frequencies. Following the NOE mixing period, the second HSQC occurs wherein the INEPT ^{13}C π pulses cover only the methyl carbon frequencies. For this purpose a 300μs Q5 pulse was used with the inversion profile given in Figure 4.9.

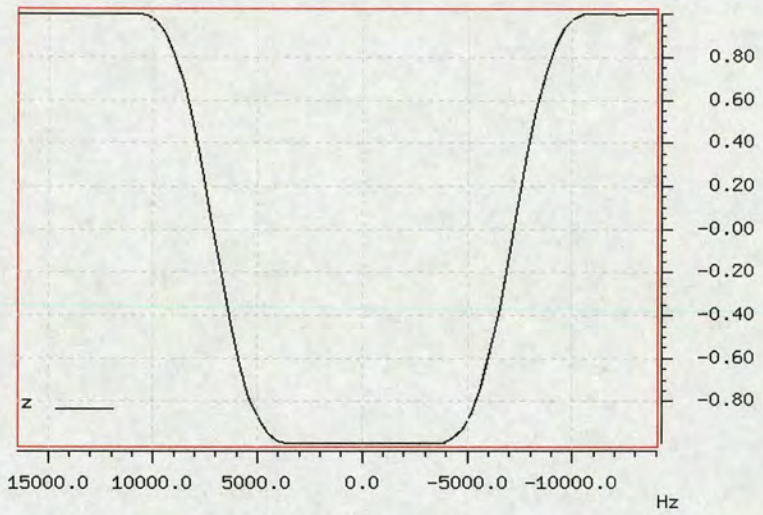


Figure 4.9: Inversion profile of the ^{13}C π pulse for selective excitation of methyl carbons

Chemical shift labeling of methyl carbon frequencies in t_2 is achieved in a constant time manner during a 28 ms interval. This is possible due to the better relaxation properties of methyl carbons. During the refocusing INEPT step, a WATERGATE block, incorporating a composite ^1H pulse is used to remove the water signal.

The product operator description of this experiment is identical to that of the CCH-NOESY with the exception of the second labeling period which, as well as being constant-time, is modified in order to edit the spectrum to remove the signals due to $^{13}CH_2$ carbons. To achieve this, once every two scans the ^1H π pulse applied during t_2 is shifted towards the start of the sequence by $1/2x^1J_{CH3}$ (ca. 4 ms). In those scans where

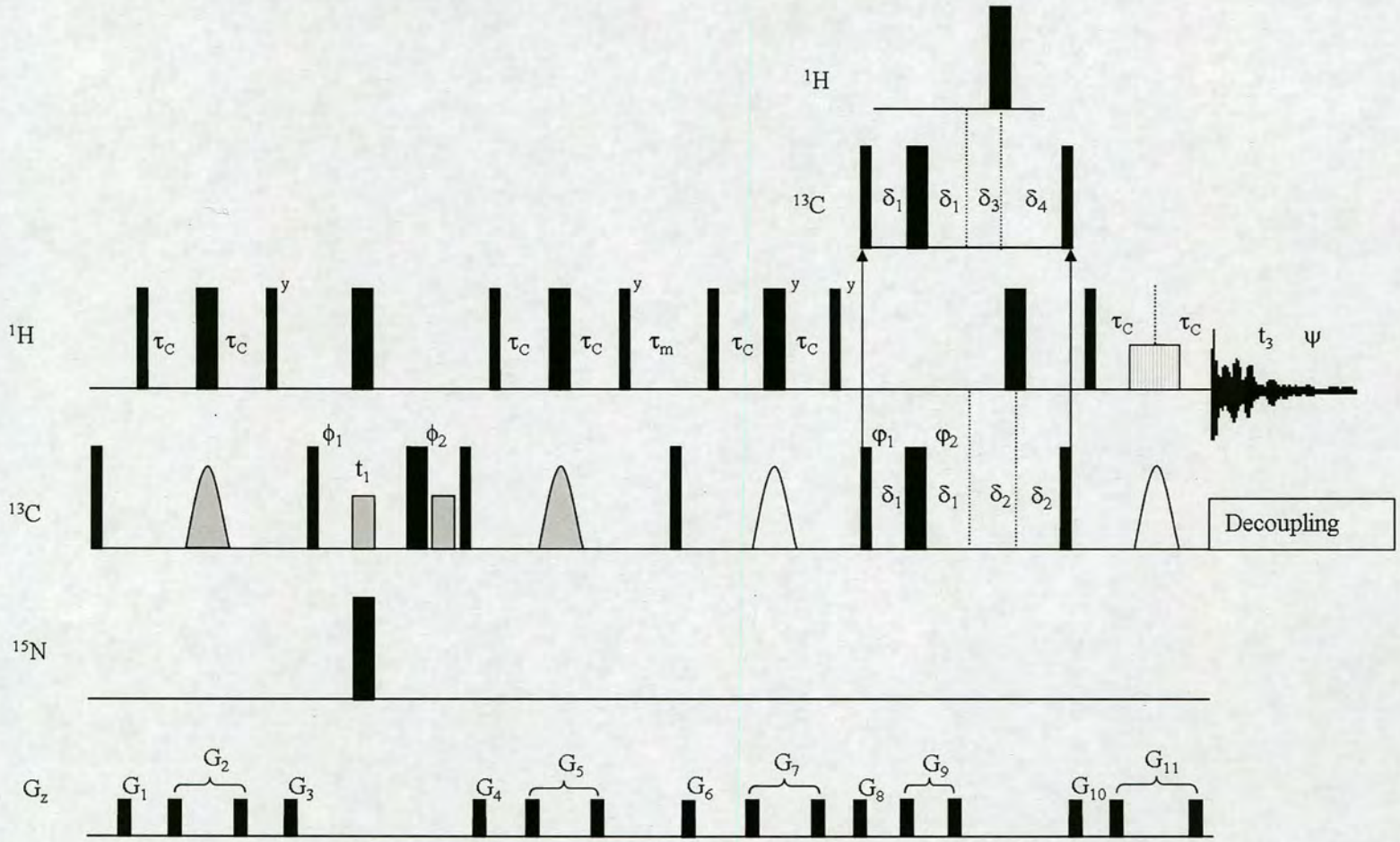


Figure 4.8: The 3D $^{13}C,^{13}CH_3$-HSQC-NOESY-HSQC sequence. $\pi/2$ and π pulses are re represented by thin and thick black bars respectively. Every two scans, the alternate sequence shown in the inset was used. ϕ_1=4x, 4-x; ϕ_2=4y,4-y; φ_1=2x,2-x; φ_2= x; ψ = x, -x, -x, x, -x, x, x, -x. All other pulses are from x axis. τ_C=1/(4x$^1J_{CH}$), τ_m=mixing time. All PFGs are 1ms except G9 (0.5ms) and applied with the following strengths: G_1=12, G_2=17, G_3=23, G_4=-14, G_5=5, G_6=18, G_7=21, G_8=50, G_9=7, G_{10}=-50, G_{11}=73. Sign discrimination in F_2 is achieved using the STATES TPPI method by incrementing φ_1 and φ_2 by 90°. Grey bars are 190μs Q5 pulses for selective ^{13}CO inversion; grey and white shaped pulses are selective aliphatic and methyl carbon inversion pulses respectively (see text). δ_1= (T-t_2)/2, δ_2= t_2/2, δ_3= δ_2-τ_a, δ_4= δ_2+τ_a, where τ_a= 1/(2 x $^1J_{CH3}$)

For product-operator treatment, see text

the ^{1}H π pulse is not shifted, no evolution of the ^{13}C-^{1}H coupling occurs and the magnetization on CH, CH_2 and CH_3 carbons will all have the same sign. Shifting the ^{1}H π pulse allows the ^{13}C-^{1}H couplings to evolve for a period equal to twice the size of the shift. At the end of this period, the effective magnetization for CH, CH2 and CH3 carbons will contain $\cos\pi J\tau$, $\cos^2\pi J\tau$ and $\cos^3\pi J\tau$ terms respectively. By offsetting the ^{1}H π pulse by $1/(2x^{1}J_{CH})$, the period of evolution will be $1/^{1}J_{CH}$ and the above terms will become -1, +1 and -1. Thus, a change in the sign of the $^{13}CH_3$ (and ^{13}CH) resonances has been achieved but the $^{13}CH_2$ resonances remain positive throughout. Suitable phase cycling of the receiver thus ensures that the $^{13}CH_2$ signals are eliminated. This removes the possibility of overlap between the $^{13}CH_2$ and $^{13}CH_3$ signals. Of course, some ^{13}CH resonances will have ^{13}C chemical shifts in the methyl region (e.g. LeuH$^{\gamma}$) but, due to their faster relaxation, these signals will have much lower intensity.

As shown in Figure 4.10, excellent resolution of methyl signals is achieved in the F_2 dimension. The spectrum is identical to the 3D CCH-NOESY spectrum with the exception that it contains only those signals that passed through $^{13}CH_3$ or ^{13}CH protons which are within the inversion bands of the selective pulses.

Due to the use of a constant-time labelling period, both the CCH_3-NOESY and its partner, the HCH_3–NOESY are less sensitive than their non-selective counterparts which show more signals and higher intensities as illustrated in Figure 4.11. However, as can also be seen, the reduced sweep width and longer sampling time in F_2 give much improved resolution and, along with the editing out of $^{13}CH_2$ signals, reduced overlap revealing more cross peaks that would have remained hidden or unresolved in the non-selective experiments. This is particularly evident in the region close to the diagonal and therefore assists identification of signals that are observed close to the diagonal in the HCH-NOESY.

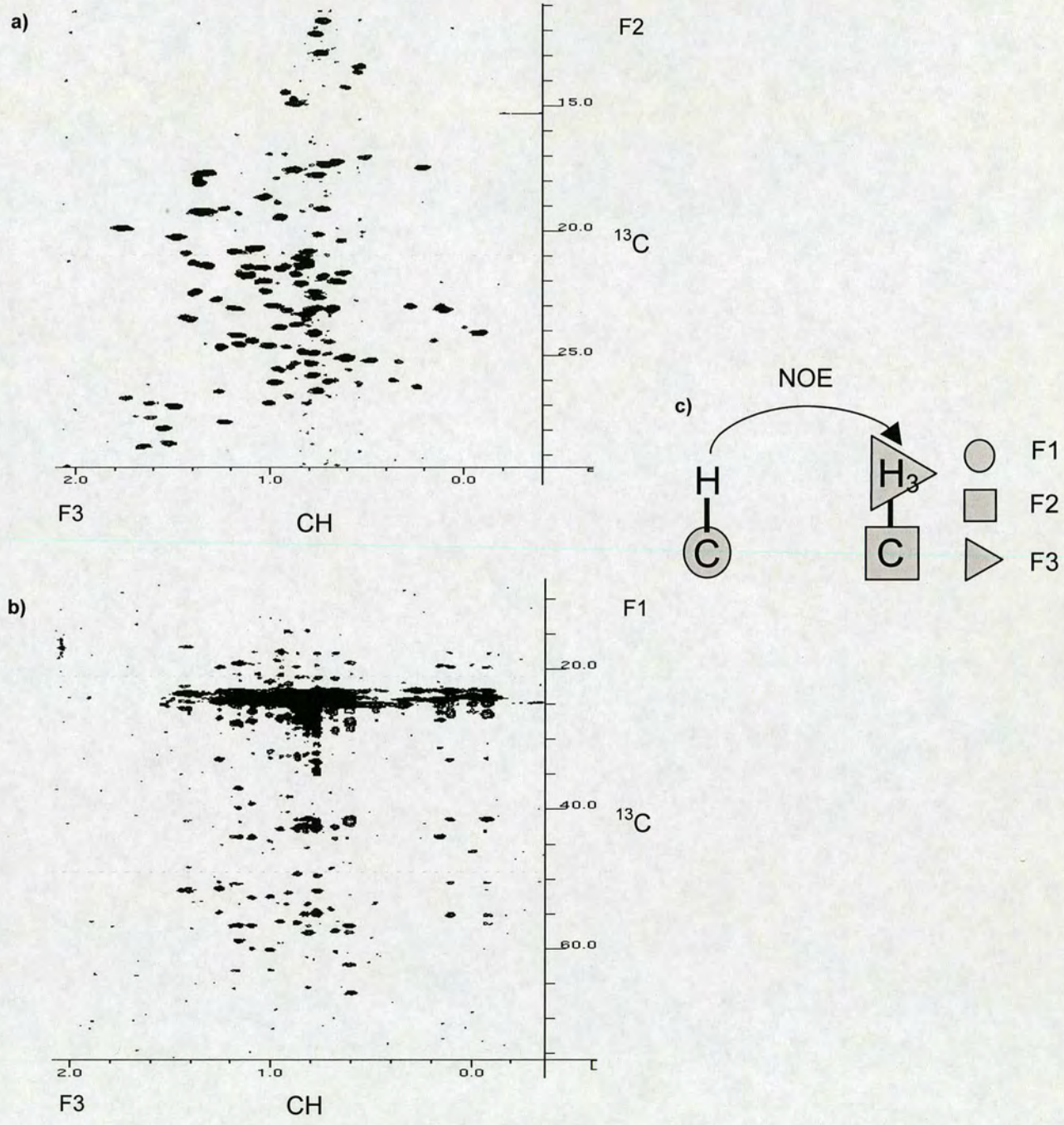


Figure 4.10: a) F_2F_3 and, b), F_1F_3 planes of the 3D CCH_3-NOESY spectrum of BlgB. c) A schematic illustration of the experiment

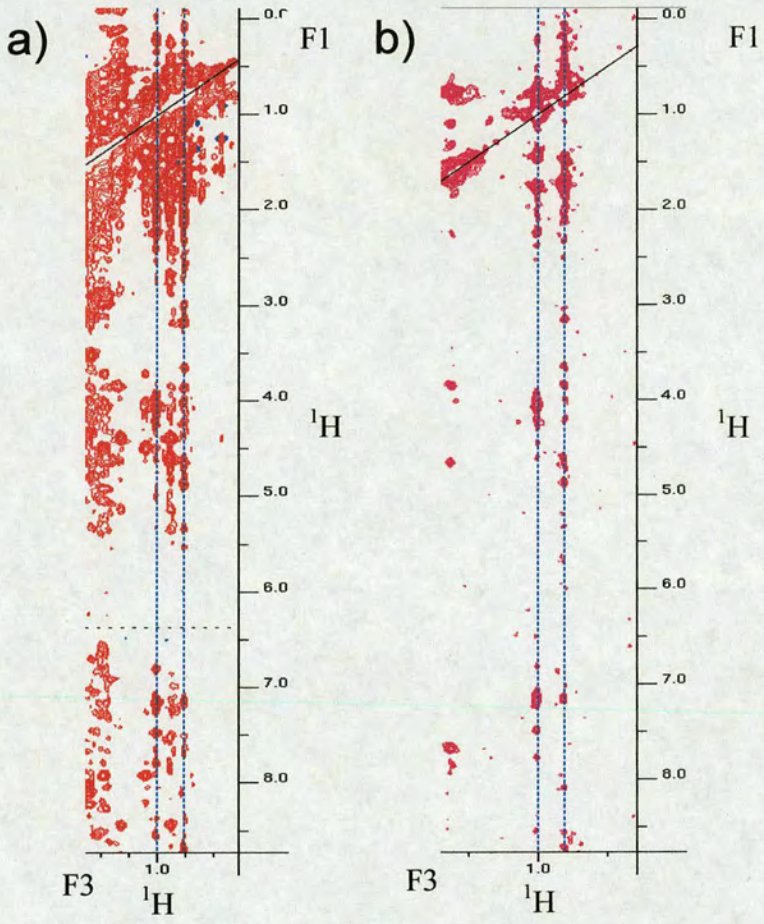


Figure 4.11: Equivalent planes of a) the HCH-NOESY and b) HCH_3-NOESY spectra of BlgB.

Chapter 5

Protocol

A total of seven 3D NOESY spectra were acquired using the BlgB sample. The development of a protocol for the use of these spectra is the focus of this chapter. Combining these spectra in an efficient and effective manner and attempting to identify those which yield the most useful information most quickly was the focus throughout this process.

5.1 Genesis of the Protocol

Using these spectra together requires some consideration so as to make the whole process as simple as possible, while maintaining the maximum assignment validity. It was decided that the basic operating principle should be the assignment of those spectra with proton-frequency modulation in t_1 (HNH-NOESY, HCH-NOESY and HCH_3-NOESY), which will be termed 'standard' NOESY experiments with the spectra with heteronuclear-frequency labeling in t_1 or 'complementary' experiments being used as assignment aids. This is the natural choice to ensure the simplicity and ease of use of the protocol; the use of the former set of experiments in structure determination via NMR being well established.

The basic principle of this method was introduced in section 2.4.4 and is summarised again here. For any strip at a particular F_2F_3 position in a standard 3D NOESY spectrum, corresponding to a particular proton, A, there will exist an equivalent strip in the

relevant complementary spectrum. In the standard NOESY strip, protons which share NOEs with A will appear along the F_1 axis at their respective chemical shifts. In the complementary NOESY strip, crosspeaks will appear at F_1 positions corresponding to the chemical shifts of heteronuclei, X, whose protons share an NOE with A. Thus, ideally, for every peak in the standard spectrum arising from an NOE interaction with an X-bound proton there will be an equivalent peak in the complementary spectrum. When considering the assignment possibilities for a peak in the standard spectrum, the probability of a correct assignment will be increased by examining the equivalent strip in the complementary spectrum. In effect, incorrect assignment possibilities of standard NOESY peaks are filtered out by comparison with the peaks in the complementary spectrum.

This is the theoretical principle of operation. However, on considering the potential for non-ideal behaviour, there emerge several points which must be addressed in constructing the eventual assignment protocol. The various cases are given below.

Case A

The simplest and most favourable case is that the spectra are perfect and contain all the peaks which would be expected. In this ideal situation, every peak in the standard spectrum will have only one (correct) assignment which agrees with one of the peaks of the complementary spectrum.

Case B

The most obvious example of non-ideal behaviour is that the complementary spectrum may not necessarily contain the relevant peak. Due to the additional magnetization transfer step employed in the complementary experiments, these spectra will be less sensitive than their standard counterparts. This will obviously be a particular concern for very weak peaks. Furthermore, some signal loss will occur when the magnetization

resides on the heteronucleus. The degree to which this occurs will differ according to the relaxation properties of the individual heteronuclei. The combination of these two effects has the potential to significantly reduce the number of peaks present in the complementary spectrum relative to the ideal spectrum.

A qualitative assessment of this difference can be obtained by simply comparing the number of crosspeaks present in equivalent strips of the standard and complementary spectra. While this is by no means a rigorous test, a significant deviation from this state would indicate a poor signal density and beg further investigation.

Case C

i) A possibility for incorrect assignment may arise if the complementary spectrum does not contain the appropriate supporting peak but *does* contain some other peak or peaks which support some other (potentially incorrect) assignment or assignments.

ii) Another point to consider is that there may be more than one assignment possibility for which a corresponding peak can be found in the complementary spectrum. This is similar to the example given in section 2.3.4:Method 1 of finding multiple symmetry-related peaks supporting various different assignments.

Both of these cases can be addressed to a large degree by looking for symmetry-related peaks to confirm assignments (once the complementary spectrum has been used to eliminate assignment possibilities) as in Method 1. If erroneous assignments are produced via either of the above instances, then, in a perfect case, there will be no symmetry related peaks to support these assignments. Of course, the absence of a symmetry-related peak may be due to factors other than the absence of an NOE interaction and it may still be possible that an 'incorrect' symmetry peak is found. However, this additional check increases the confidence with which any assignment is made and the number of cases where both an incorrect complementary peak *and* an

incorrect symmetry peak support an assignment should be far less than the cases of incorrect complementary peaks alone. Looking for symmetry related peaks is therefore crucial to the operation of the Protocol.

Case D

A significant problem arises if the proton shift is known but the shift of the heteronucleus to which it is bound is not known. In these cases, not only is it very difficult to look for symmetry related peaks but using the complementary peaks to filter the ^{1}H possibilities will, in fact, eliminate the correct assignment.

Obviously, this must be accounted for and so a condition is added to the protocol whereby any assignment possibility for a standard NOESY peak, whose heteronuclear shift is not known, is considered to be a valid assignment thus ensuring that no potentially correct assignments are falsely disregarded.

Case E

The peaks of the standard and/or complementary spectra may be overlapped with one another and therefore cannot be picked. For the standard spectra this is not an uncommon problem and it is somewhat benign since, importantly, it does not lead to erroneous assignments, merely a loss of information. For the complementary spectra, this is a somewhat greater concern as being unable to identify an individual peak may well lead to incorrect assignments.

The operation of the protocol will therefore be as shown in Figure 5.1. Its basic feature being that assignment possibilities are filtered via the complementary spectrum prior to looking for symmetry related peaks.

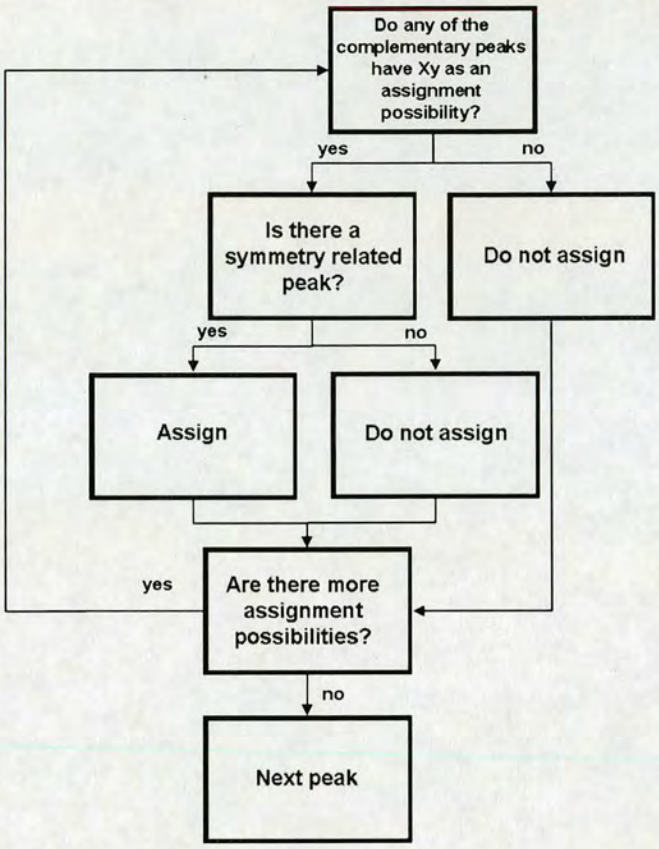


Figure 5.1: Flowchart of the decision making process used in assigning a NOESY spectrum via method 2 considering a particular assignment possibility, H^y, (where y is the position α, β..) bound to a heteronucleus, X.

The emphasis throughout the development of the protocol is reducing the total assignment ambiguity while maximizing the confidence with which any assignment is made. This is by no means an infallible process and, as has been noted, it is still possible to make erroneous assignments. However, this method is based upon the fewest assumptions and ensures that those assignments which *are* made are made only on the basis of the available data. Thus, even though it is accepted that some assignments will be incorrect, a far greater portion should be correct and, on performing a structure calculation, it should become apparent which assignments are incorrect and can be dealt with accordingly.

Of course, the resonance assignment of the protein may be incomplete which presents further potential problems. When assigning a NOESY spectrum via the standard method, an incomplete resonance assignment may well lead to peaks being assigned incorrectly since the correct assignment will not be presented as an assignment possibility. The use of the complementary spectra can be seen as a safeguard against this given that, ideally, none of the incorrect assignments will have a corresponding peak in the complementary spectrum. Making an incorrect assignment in such a situation is still possible but the risk is certainly reduced.

5.1.1 Initial Implementation of the Protocol

Manually investigating the assignment possibilities for a particular peak is a time-intensive procedure and, arguably, the addition of this intermediate stage (relative to Method 1) will only lengthen the time spent assigning the spectra.

It is therefore desirable to employ some automated or semi-automated processing of the assignment possibilities such that, for any given peak, the user is only presented with those assignment possibilities for which there exists an appropriate peak in the complementary spectrum. In so doing, the total number of possibilities which must be examined will be reduced relative to Method 1 and progress can be made more rapidly. There follows a brief explanation of the implementation of this procedure in the Analysis software package.

On assigning a peak in one of the standard NOESY spectra, the user is presented with all the nuclei whose chemical shift lies close to the chemical shift of the peak under consideration, in that dimension. In Analysis this is achieved by searching the chemical shift list for all shifts that lie within a certain, specified range of the chemical shift value. For each of the possibilities, the equivalent complementary strip must be examined to see if it contains a peak with a plausible heteronuclear chemical shift.

Naturally, the assignment process could be conducted on any peak and in any order, however, as discussed, the potential for multiple valid assignments, missing peaks or overlap mean that a logical approach would be to examine the spectra one strip at a time. In addition, since the peaks of a particular strip in the standard NOESY must all be checked against the assignment possibilities of the peaks of the same strip of the complementary spectrum, it is also more efficient to assign the spectrum in this way, thus minimizing the time spent selecting the correct regions in each spectrum and the, albeit small, computation time.

There are numerous ways to automate this process but the simplest, in terms of those functions available within Analysis, was to first create a list of all assignment possibilities for the peaks in a strip of the complementary spectrum. An assignment possibility for a peak in the standard NOESY strip can then be checked against this list to see whether there are any 'matches', indicating a valid assignment. This is achieved by querying whether the two nuclei are covalently bonded to one another.

The overall procedure for assignment of the spectrum is thus: i) select the standard NOESY strip and the equivalent complementary strip, ii) pick the peaks in each strip iii) execute the macro to yield the assignment possibilities deemed valid. On executing the macro, the graphical output shown in Figure 5.2a is generated containing numbers which correspond to the peaks of the standard NOESY strip in the order they were selected. In this example the spectra being used are the HNH-NOESY and the CNH-NOESY. Transfer of the F_2 and F_3 assignments (in this case, from the 2D ^{15}N-^{1}H HSQC) is a trivial matter. For this purpose, the 'Assign other Dims' button is used. The user can then select an individual peak from the list and on clicking the 'Assign' button, a second window is opened, as illustrated in Figure 5.2b, displaying the assignment possibilities for that peak deemed valid by matches with assignment possibilities of the complementary spectrum peaks.

a) **b)**

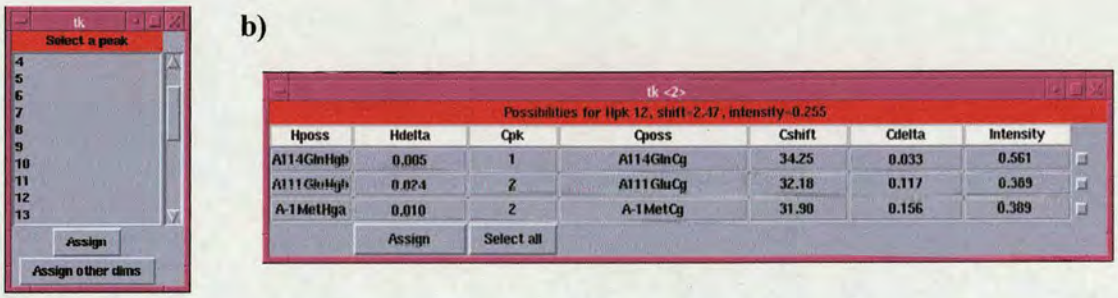


Figure 5.2: Graphical output of the semi-automated assignment macro showing, a), the peak selection window and, b), the assignment window for a particular peak

The user must then investigate each of the assignment possibilities identified to see if a symmetry related peak exists by manually navigating to the appropriate region of the relevant spectrum (in this case, the C^{ali}-H^N region of the HCH-NOESY).

The assignment possibilities window contains some further information that may potentially be of use when assigning. The first of these is the difference in chemical shift between the assignment possibility and the peak, referred to here as delta. In this procedure there are two delta values to be considered – one for the standard NOESY spectrum and one for the complementary spectrum. If both of these values are large, the assignment possibility is less likely to be true than if both assignment possibilities agree well with the positions of the peaks from which they are derived. The second item, denoted Cpk, shows which of the peaks in the complementary strip gave rise to the matching assignment, allowing the user to locate the relevant peak easily. The third item is that of relative intensity. This gives the intensity of a complementary spectrum peak relative to the most intense peak in that strip (which is given an intensity of 1). This can be compared with the relative intensity of the standard spectrum peak (given at the top of the window) to provide a further assessment of the validity of a particular assignment. These intensities are not expected to match exactly for correct assignments due to overlap and relaxation considerations, nevertheless, a large difference will indicate an assignment is less likely. A further improvement was made to this interface to indicate

when a particular peak of the complementary spectrum had already been 'used'. This is an attempt to account for the fact that, in the absence of overlap, a single peak in the complementary strip should correspond to no more than two peaks (in the case of non-equivalent methylene protons) in the standard strip. If a 'match' is made then the complementary peak involved in that match is flagged (coloured red in the Cpk column) letting the user know it has been used. This provides a further means of eliminating assignment possibilities. The relative intensity values are also of use here to indicate which peaks show good agreement in intensity and are therefore likely partners or which complementary peaks can potentially be used more than once, reflecting the possibility of overlap.

The assignments which are deemed valid after checking for symmetry peaks, comparing intensities and checking for peaks which have been used more than once can then be selected via the radio buttons at the right of each row and assigned to the peak using the 'Assign' button.

5.1.2 Initial Testing of the Protocol

Prior to testing the performance of the assignment protocol, a qualitative measure of the suitability of the HNH- and CNH-NOESY spectra of BlgB was obtained by comparing the number of peaks present in 30 equivalent strips of the two spectra. The results are given in Table 5.1

Spectrum	No. peaks
HNH-NOESY	449[a]
CNH-NOESY	313

Table 5.1: Numbers of peaks present in equivalent strips of the HNH-NOESY and CNH-NOESY spectra. [a]This is the number of peaks present in the aliphatic ^{1}H region of each strip.

Each peak in a strip of the CNH-NOESY spectrum should correspond to one (in the case of CH, equivalent CH_2s and CH_3 groups) or two peaks (non-equivalent CH_2s) in the HNH-NOESY strip and so the number of peaks present in each should be fairly similar. In this case, the number of peak present in the complementary spectrum relative to the standard spectrum was ca. 70%. This was judged to be a suitable agreement to proceed with using these spectra.

The procedure outlined above was then used to assign the HNH-NOESY spectrum of BlgB. In order to assess the performance of the protocol relative to the standard assignment method, two separate, identical peak lists were created for the spectrum to allow each peak assignment to be performed twice - once using Method 1 and once using Method 2. This affords a direct, side-by-side comparison of the two methods to examine not only the degree to which ambiguity has been reduced but also what impact this difference has upon the subsequent structure calculation. The results of this process are summarized here. Of course, when using the CNH-NOESY spectrum, only the peaks corresponding to H^N-H^{ali} interactions can be assigned via Method 2, the remaining peaks (H^N-H^N and H^N-H^{aro}) were assigned via Method 1.

On considering the results of assignment via method 2 compared with those of method 1, two situations are immediately apparent where the method 2 outperforms method 1. The first of these cases is that of ambiguity elimination. Here the protocol returns a single, unambiguous assignment for a peak while the standard method produces an ambiguous assignment. Figure 5.3 gives an example of this. In this particular case, the assignment provided by the protocol was confirmed by the presence of a symmetry related peak in the HCH-NOESY spectrum and was therefore assigned to the peak. Of the standard method assignments, it was not possible to exclude any possibilities on the basis of symmetry related peaks and, thus, the standard method would require that all three possibilities be assigned to the peak.

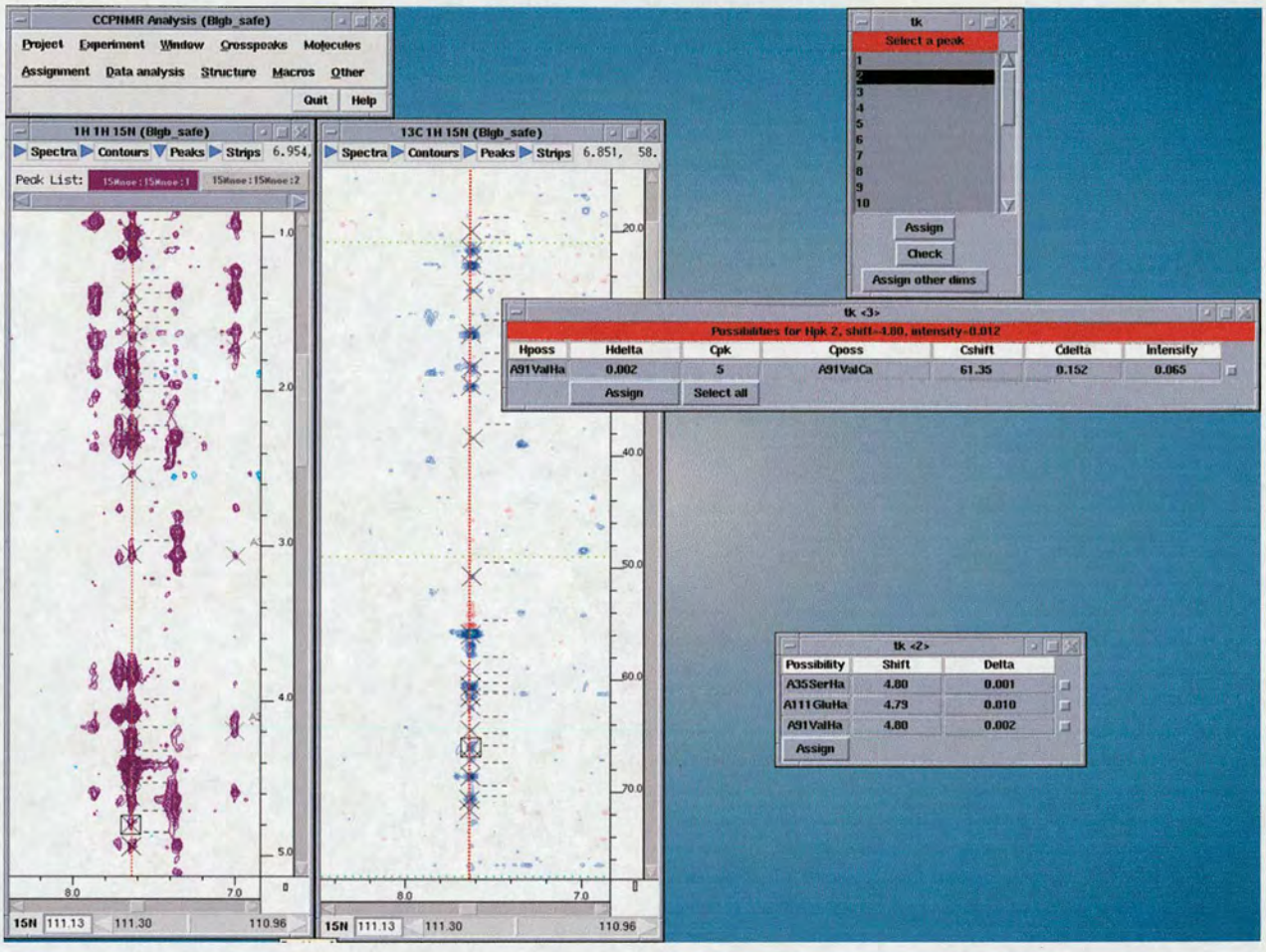


Figure 5.3: Screenshot of semi-automated assignment macro in operation within the Analysis software package. Illustrated is a case of ambiguity elimination. The peak under consideration (highlighted) has three valid assignment possibilities when assigned using the standard method (bottom right window) but only one possibility when assigned via the protocol

The second instance is that of ambiguity reduction. In this case both the protocol and standard method assignments are ambiguous but the degree of ambiguity is less for the protocol than the standard method. An example of this is shown in Figure 5.4.

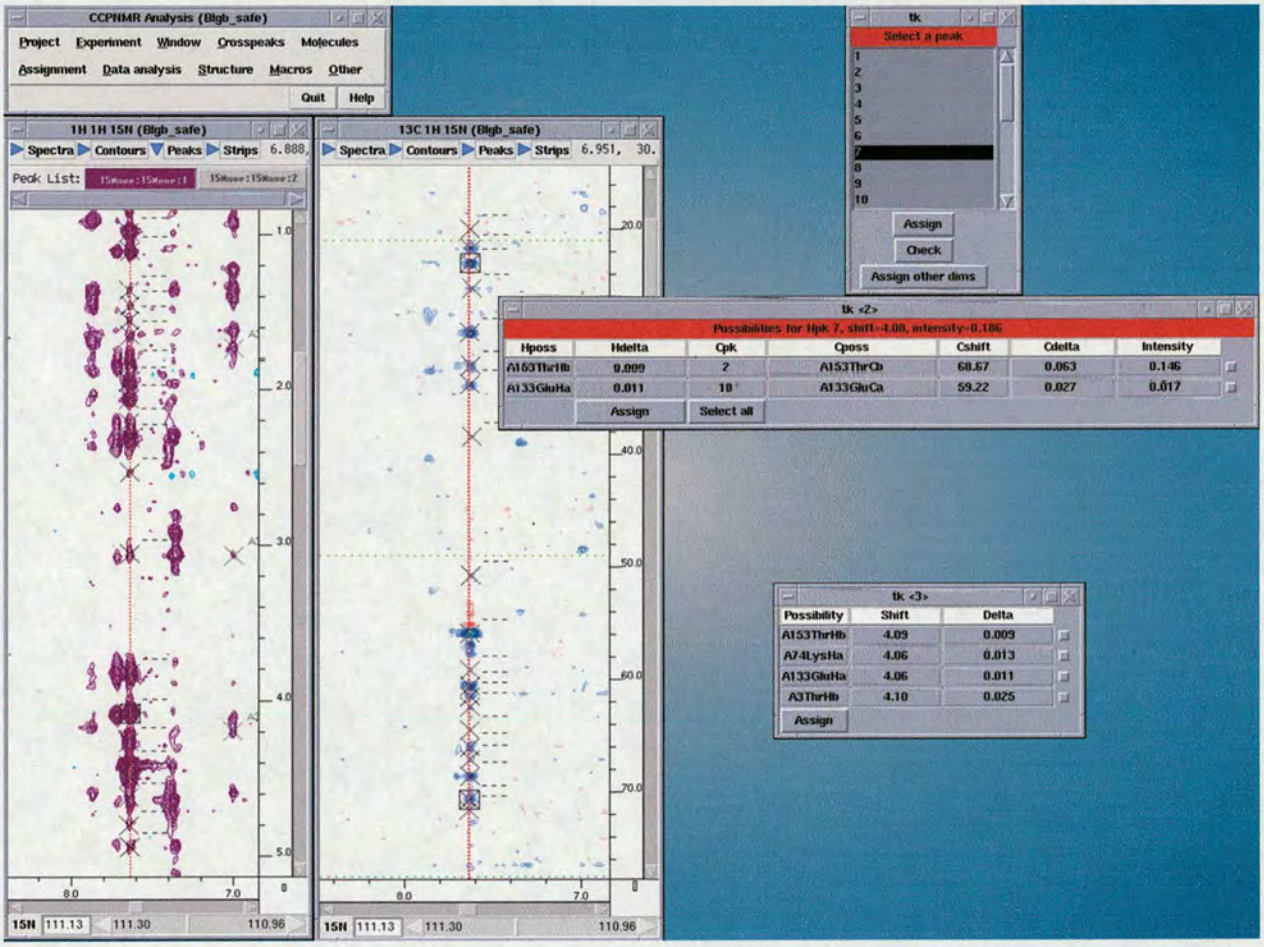


Figure 5.4: Screenshot of semi-automated assignment macro in operation within the Analysis software package. Illustrated is a case of ambiguity reduction. The peak under consideration (highlighted) has four valid assignment possibilities when assigned using the standard method (bottom right window) but only two possibilities when assigned via the protocol (neither of which can be eliminated on the basis of symmetry related peaks)

During this process of testing, a third potential improvement emerged. As explained previously, the semi-automated assignment macro tracks which of the peaks in the complementary spectrum strip have been used to corroborate assignments in the peaks of the standard spectrum. In some cases, upon completing the assignment of a standard spectrum strip, it was noted that some peaks of the complementary spectrum remained

unused. Clearly these peaks must have arisen due to some NOE interaction and should therefore have a corresponding peak in the standard strip. On examining the assignment possibilities for these unused peaks to determine where the corresponding peaks should be in the standard spectrum, it was found that, in the majority of cases, the predicted peak positions were in crowded regions of the standard spectrum strip. Of those few cases where no potential partners were found in the standard spectrum, the complementary spectrum peaks under consideration were relatively weak and were eventually regarded as noise.

Thus, the protocol also provides a method for extracting additional information from the standard spectra that would otherwise have remained hidden. Similar observations have been made previously (Diercks et al., 1999). An additional function was therefore included in the macro to indicate which complementary peaks had not been used. This information is obtained using the 'Check' button visible in the previous figures.

Recording this information can be achieved in two ways – the 'unused' complementary peak can be assigned to the appropriate ^{13}C nucleus in F_1 or a new assignment can be created at the appropriate position in the standard spectrum and assigned accordingly. The former approach is somewhat less accurate since it would not indicate those cases where only one of the protons bound to a heteronucleus shows an NOE to the proton under consideration and therefore the latter approach was adopted. A problem arises, however, in attempting to convert these instances into useable restraints for a structure calculation. Since the new assignments which are created will exist in regions of intense signal, translating the signal intensity into a particular distance class may well lead to these restraints being incorrectly classified. To account for this, it was decided that all such restraints should be placed in the longest (0-6Å) bound.

Overall, the peaks were classified in three categories – unambiguous, semi-ambiguous and ambiguous, reflecting the number of assignments and the degree of certainty with which they were made. The three classes are defined as follows:

• Unambiguous: F_1, F_2, F_3 assignments known

• Semi-ambiguous: Multiple F_1 assignments possible

• Ambiguous: No F_1 assignment made

The two types of ambiguity differ in the amount of information supporting their assignments and are therefore treated differently when used in the structure calculation.

At this point it is useful to consider the order in which the various restraints (unambiguous, semi-ambiguous, ambiguous) are used in the calculation procedure. If incorrect assignments exist, the calculation process is relied upon to detect their presence. This requires that the incorrect assignments be outweighed by a sufficient majority of correct assignments to ensure that the incorrect assignment cannot be satisfied whilst satisfying all other restraints. This may become more difficult if ambiguous assignments are included at the start of the process. Since an ambiguous assignment provides a number of potential satisfactory outcomes it is possible that incorrect assignments may be 'masked' by the greater flexibility or range of calculated structures. It was therefore deemed preferable to proceed with unambiguous restraints as far as possible before introducing ambiguity into the procedure. The weight of correct unambiguous restraints will thus allow identification of the incorrect restraints (as violations). These can then be dealt with and the first generation of structures will be produced. Ideally, this ensemble of structures will be of sufficient quality to be used to filter the ambiguous restraints, thus decreasing the time which must be spent dealing with these restraints. In addition, the quality of the structure ensemble which is used in the filtering process will determine how effective the process is – the more accurate the ensemble, the more incorrect contributions to ambiguous assignments will be discarded.

Inclusion of the ambiguous restraints is performed in two stages with the completely ambiguous restraints the last to be added. This is due to the assignments given to these peaks being made solely on the basis of chemical shift (no supporting complementary peak or symmetry-related peaks being found) and, therefore, these assignments are supported by the least information. That no symmetry related peaks were found could, among other things, be due to these peaks being items of noise or artefacts or it may be the result of incomplete chemical shift tables failing to provide the correct assignment possibility. Therefore, there is a greater likelihood that these restraints contain erroneous information which would hinder the process of structure determination if included earlier in the proceedings.

The assignments made based on extra signals found in complementary spectra are also somewhat less certain – the existence of a peak in the standard NOESY being inferred rather than explicit – and the intensity class, as mentioned, is also not precisely known. Also, the total number of such cases was found to be quite few (27) and the majority of these were the less significant intra-residue and sequential NOEs. For these reasons it was also decided to defer the inclusion of these restraints until a later point in the calculation.

In total, 2913 peaks were selected. The number of unambiguous, semi and ambiguous assignments obtained for Methods 1 and 2 are given in Table 5.2.

Method	Unambiguous	Semi-ambiguous	Ambiguous
1	1749	742	422
2	1867	607	439

Table 5.2: Numbers of unambiguous and ambiguous assignments produced assigning HNH-NOESY spectrum via Method 1 and Method 2

From this, it is clear that Method 2 gives a greater number of unambiguous assignments and, although the magnitude of this increase is perhaps smaller than expected, the significance of these additional restraints will soon become clear.

It should also be noted that these figures are the initial assignments, prior to any structure calculation and subsequent investigation of violated restraints and will therefore be expected to contain a number of 'incorrect' assignments. Furthermore, these numbers are for all the peaks in the spectrum, corresponding to all interactions - H^N-H^N, H^N-H^{aro} and H^N-H^{ali}. Using the CNH-NOESY spectrum will only reduce the ambiguity of the H^N-H^{ali} peaks, however which comprise 1360 of the Method 1 unambiguous assignments and 1474 of the Method 2 unambiguous assignments.

Thus, the protocol performs as expected by reducing the total assignment ambiguity, but whether this has a significant effect upon the subsequent structure calculation is a far more important consideration. To this end, the unambiguous peak assignments were converted into restraints using the Generate Constraints command in Analysis. The resulting restraint tables were exported and used as input in a CNS structure calculation.

Following these calculations, the restraints which were violated were examined. As noted earlier, when using Method 2, an incorrect assignment can be made due to some lack of information i.e. the correct complementary peak was not present but some other, incorrect assignment *was* supported (case C). These instances cannot be detected, however, until the structure calculation is unable to satisfy these restraints and they are presented as violations. Checking these violations therefore involves performing assignment of the relevant peaks via Method 1 in the hope of obtaining the correct assignment. If assignment via Method 1 reveals no more viable assignments, the peak is made ambiguous.

After checking the violations there remained 1673 unique (non-duplicate), unambiguous restraints in the Method 2 dataset and 1527 unique, unambiguous restraints in the Method 1 dataset. At this point, the structures calculated with these restraint tables began to show the correct fold as shown in Figure 5.5.

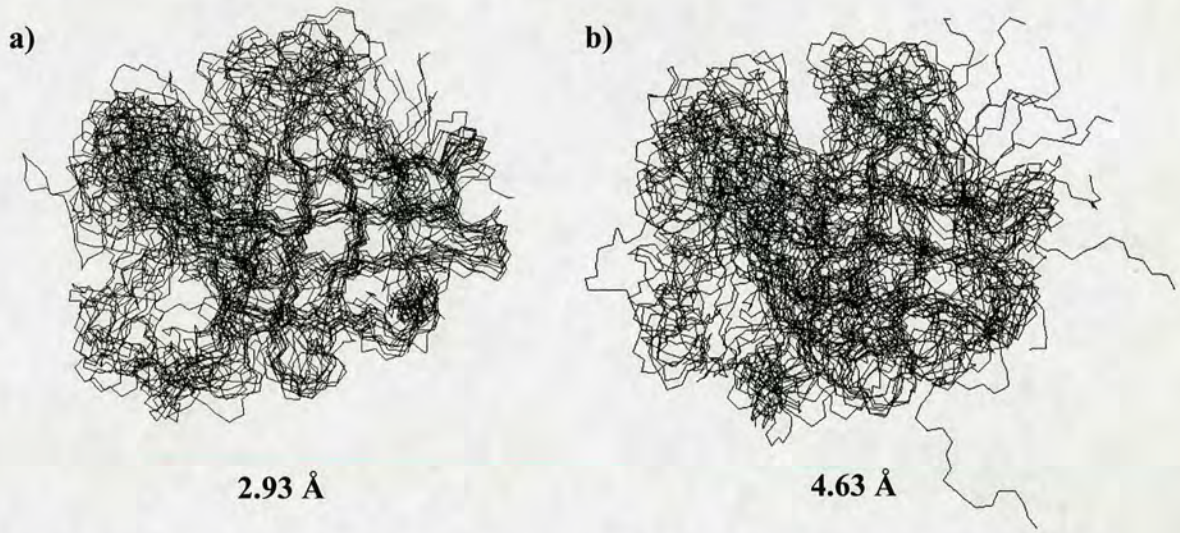


Figure 5.5: Lowest energy (20 of 100) structures resulting from structure calculations performed using a) Protocol restraints and, b), Method 1 restraints. The numbers indicate the backbone rmsd's. Figure made using MOLMOL (Koradi et al., 1996)

The structures resulting from the Method 2 restraints are seen to be superior to those obtained using the standard method restraints with lower rmsds and, significantly, structures converging on the same overall fold. This is not the case for the structures obtained using Method 1 assignments where no two structures enjoy the same overall fold, although, as can be seen, elements of structure are evident. Thus, it would appear that Method 2 reduces the total assignment ambiguity and, in doing so, provides a set of restraint data which deliver better structures more quickly.

The number of restraints used in each calculation does not differ greatly and so the differences observed here must be due to the type of restraints present in the two datasets. The NOE statistics for these unambiguous restraints are given in Table 5.3.

	NOE type			
Method	**Intra-residue**	**Sequential**	**Short Range**	**Long Range**
1	471	717	142	197
2	471	716	196	290

Table 5.3: NOE statistics for unambiguous restraints produced via Method 1 and 2. Short range restraints correspond to interactions between nuclei which are 2-4 residues apart in the sequence and long-range to nuclei >4 residues apart.

As would be expected, the majority of the restraints are intra-residue and sequential restraints and these are essentially the same for both methods. The significant difference is found in the number of short and long-range distance restraints which have a much greater effect upon the structure determination process. Method 2 is seen to produce ca. 35% more short range and ca. 47% more long range distance restraints relative to Method 1.

The distribution of assignments reflects a natural conservatism employed in assigning NOESY spectra whereby the strongest NOE peaks usually report intra-residue and sequential contacts. Such peaks are therefore assigned unambiguously more frequently than the weaker short- and long- range crosspeaks.

In a very few cases it was noted that where Method 2 returned no assignment (ambiguous) Method 1 in fact returned an unambiguous assignment. These instances are due to occurrences of case B i.e. the CNH-NOESY spectrum did not contain the relevant peak (or the region was too crowded to pick the relevant peak). This results in the Method 1 dataset containing some additional unambiguous assignments which are not present in the Method 2 dataset. In total, nine such cases were observed – six sequential NOEs and three short-range NOEs.

The significance of the reduced ambiguity inherent in Method 2 is supported by QUEEN analysis of the unambiguous restraints provided by Method 2. Figure 5.6 shows the results of this analysis. Indicated on this plot are those assignments which were obtained via Method 2 but *not* via Method 1 i.e. the cases of ambiguity elimination. In total, 155 ambiguity elimination events were identified representing 11% of the unambiguous H^N-H^{ali} crosspeaks. The NOE statistics for these ambiguity elimination events are given in Table 5.4. Of these 155 ambiguity eliminations, the majority of these can be seen to be the more structurally significant short- and long-range restraints.

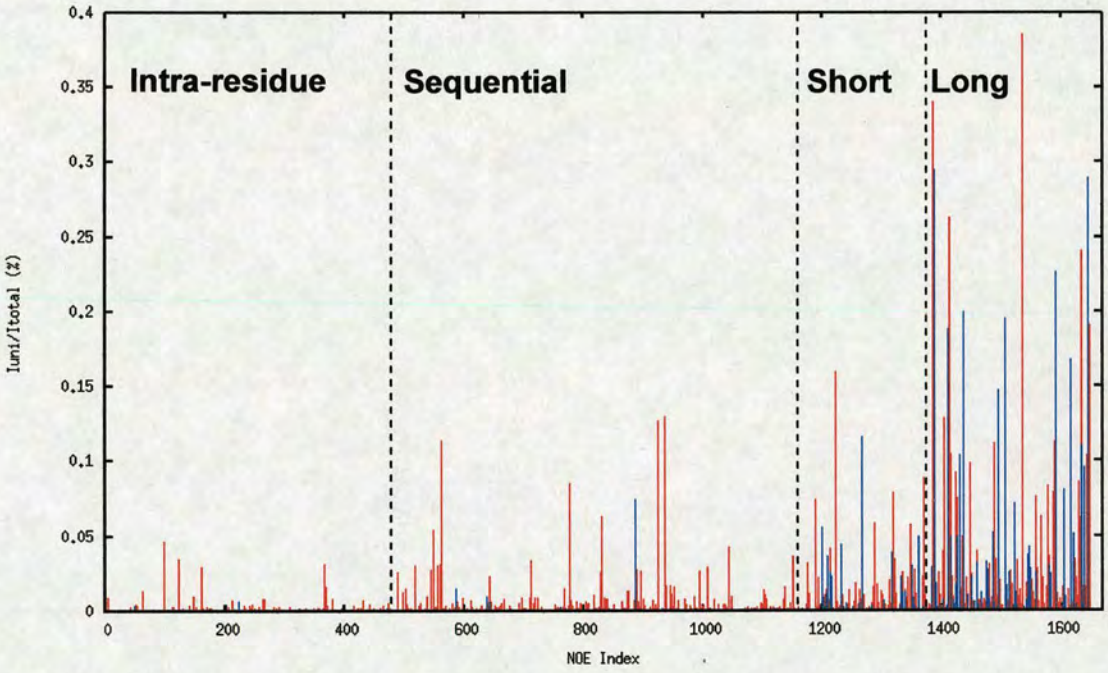


Figure 5.6: Plot of I_{uni}/I_{total} values for unambiguous restraints obtained via Method 2. Restraints which correspond to ambiguity elimination events are in blue.

NOE class		
Sequential	**Short**	**Long**
5	57	93

Table 5.4: Numbers of sequential, short- and long-range restraints present in the 155 cases of ambiguity elimination.

It is clear that Method 2 provides more unique information which leads to the correct structure for the protein more quickly.

Examining the semi-ambiguous restraints shows a similar result. A peak by peak comparison of the degree of ambiguity (number of assignments) in the semi-ambiguous Method 1 and 2 restraints was performed to obtain the number of cases of ambiguity reduction. Again, the total ambiguity which exists is reduced with a total of 207 cases of ambiguity reduction being identified (ca. 14% of the H^N-H^{ali} restraints).

The effect of an improved structural template and restraints with reduced ambiguity as input for the filtering process is shown in Figure 5.7. Proceeding via the Protocol gives a better quality structure much more quickly, arguably requiring only one round of filtering. Method 1, however, requires more filtering rounds to achieve the same quality of structures.

The assignment of the HNH-NOESY took several months. Method 1 and Method 2 were conducted in tandem and so an assessment of the time difference in assigning the spectrum by each method was not possible. The steps involved in each method are essentially the same, however, the only difference being the number of possibilities which must be considered for each peak (Method 1 requiring more). Method 2 therefore did not provide a particularly significant time gain in this area, though it certainly requires no more time than Method 1. More significant was the time required to derive a structure. In the early calculation rounds, prior to the filtering process, Method 2 quickly began to show the correct fold after only two rounds. By contrast, the structures derived from the Method 1 restraints did not show similar results until the fifth round. This can be attributed to the increased amount of information present in the protocol restraints (due to the total reduction of ambiguity) leading to a *decreased* proportion of *incorrect* information. The Method 1 dataset contained a greater number of incorrect restraints than the Method 2 dataset and, as has been shown, the Method 2 dataset had more structurally significant information.

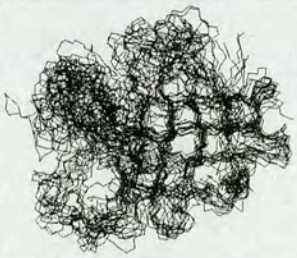
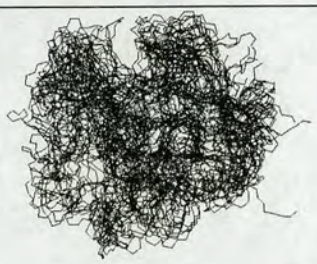
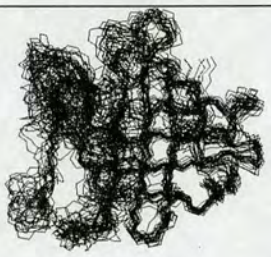
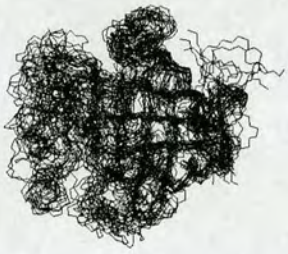
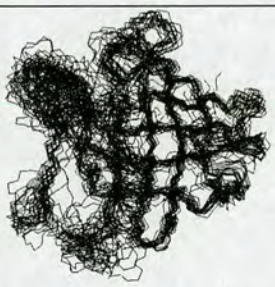
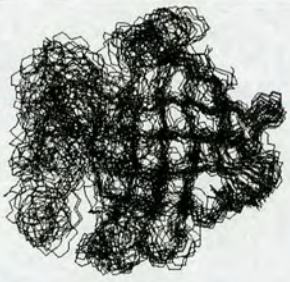
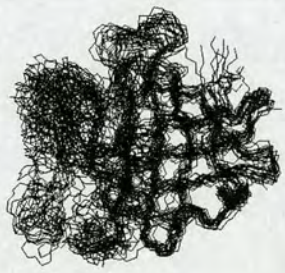
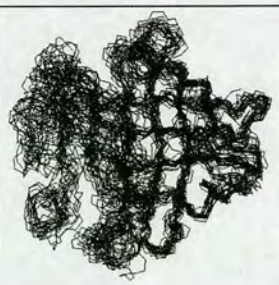
Round	Method 2	Rmsd (Å)	Method 1	Rmsd (Å)
1		2.93		4.63
2		2.09		3.56
3		1.97		2.77
4	----------			2.20
5	----------			2.02

Figure 5.7: Comparison of filtering steps performed using Method 1 and 2 assignments

5.1.3 Streamlining the Protocol

Comparing the structure calculation results of both methods show that Method 2 clearly outperforms Method 1, that is, it provides more, correct unambiguous information, sufficient to derive a better initial structure. The time taken in assigning the HNH-NOESY spectrum was quite considerable, however, and thus, prior to the assignment of the remaining spectra, an alternative, less detailed and faster assignment method was tested to see whether the total assignment time could be substantially reduced without a significant increase in incorrect or ambiguous assignment.

At its core, the assignment protocol for any peak consists of finding those assignment possibilities deemed valid by the existence of an appropriate peak in the complementary spectrum and then confirming or eliminating them by looking for symmetry related peaks. The first step (filtering assignment possibilities via the complementary spectrum) is already automated, and has been shown to produce valid assignments and the user is required only to search for symmetry related peaks. In the method described previously, this was achieved manually and a number of additional items were used to aid in choosing the correct assignment. A reduced interaction assignment protocol is possible whereby the user is only required to confirm or deny the existence of a symmetry related peak.

The first difference in this new scheme was the manner in which peak were submitted to the macro. With the HNH-NOESY and CNH-NOESY spectra both having been picked, it was decided that individually selecting the strips of peaks in these spectra to submit to the macro was unnecessary and, instead, the macro creates 'virtual' strips by grouping the peak objects together within Analysis according to their F_2 and F_3 assignments. The macro then automatically filters the assignment possibilities for a peak via the complementary spectrum. For each of the resulting possibilities, the user is automatically presented with the region of the appropriate spectrum corresponding to the expected location of a symmetry related peak. The user is then asked whether a

symmetry related peak exists and an interactive GUI query is displayed to receive the result of this decision. If it is deemed that a symmetry peak exists, the peak is assigned appropriately. As previously, if a peak has no matches in the complementary spectrum or no symmetry peaks are found, the peak remains unassigned. An example of this new macro in operation is given in Figure 5.8. In this example, the complementary spectrum is not shown – it is incorporated as a filter 'behind the scenes'.

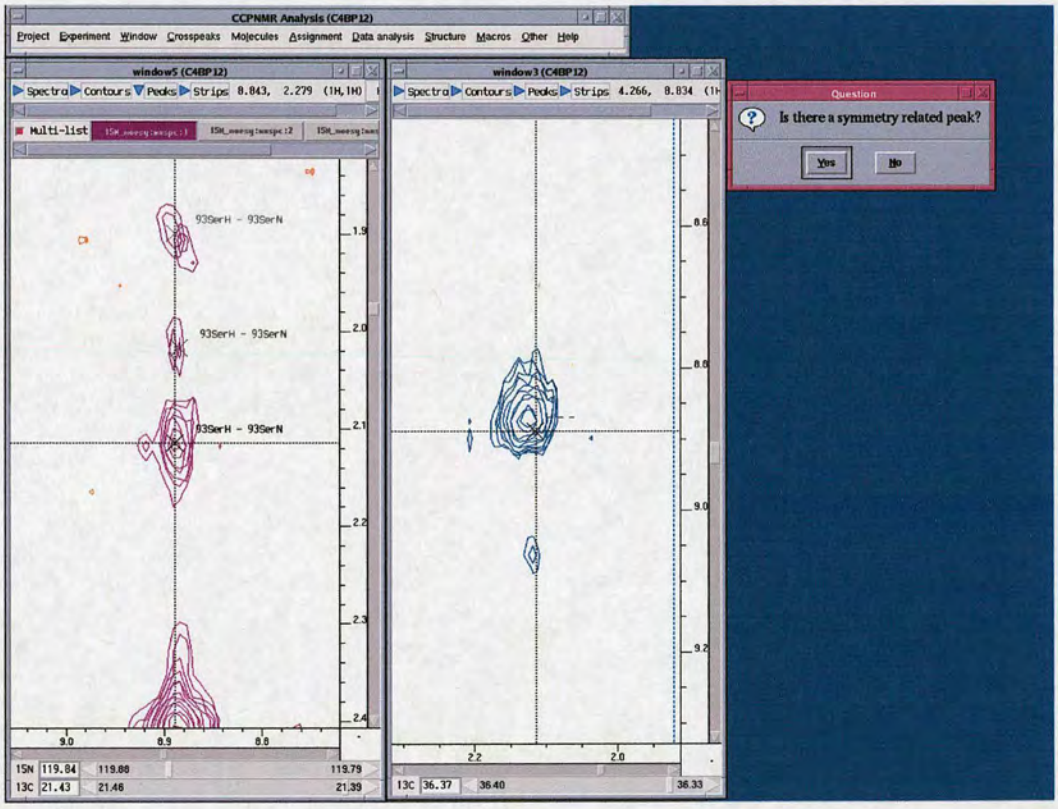


Figure 5.8: Screenshot showing the streamlined macro in operation. The HNH-NOESY peak being assigned is displayed (leftmost window) along with the appropriate region of the HCH-NOESY spectrum (center window), containing the expected symmetry related peak. The dialogue (rightmost window) is used to confirm whether the assignment possibility is supported by a symmetry related peak or not. In this case, a symmetry peak is present and the assignment is made.

Using this method, the user need not select which peaks are to be assigned (and used as assignment aids), the macro automatically groups the appropriate peaks together. Individually selecting the peaks to be used is still possible, however. Once a peak has been assigned (or examined but left unassigned) by the protocol, it is marked as such using the 'details' attribute of the peak to allow the macro to ignore it in future. The significant time reduction of this process comes from eliminating large parts of manual interaction – selecting the peaks to submitted to the macro, selecting the peak to be assigned, navigating to the expected symmetry peak location, deciding whether an assignment is valid on the basis of chemical shifts, intensities and symmetry related peaks. By reducing the procedure to its most basic components and employing automation to a large extent, this streamlined procedure results in far less time being spent on the assignment process – several strips can be assigned in the time taken to assign just a few peaks.

In order to test the performance of this alternative method, the assignment of the HNH-NOESY spectrum was repeated using the new macro. Only the peaks due to NOEs with aliphatic protons were reassigned since the H^N-H^N and H^N-H^{aro} peaks were assigned using the standard method and can therefore be assumed to remain the same. On testing this mode of operation on individual strips of the HNH-NOESY spectrum, it was found that, in general, the less detailed, faster method, gave assignments which were identical to those of the original, semi-automated method. In those few cases where the assignments were found to differ, this difference was due to allowing a peak of the complementary spectrum to contribute to more than one peak in the equivalent strip of the HNH-NOESY (and the existence of a supporting symmetry peak), thus resulting in a more ambiguous assignment for some peaks. This is a consequence of working without the additional assignment aids which were used in the initial implementation of Method 2 (e.g. used peaks, relative peak intensities). Following this second assignment, the two sets of assignments were compared to determine to what extent the two procedures differed in their degree of ambiguity. These results are given in Table 5.5.

Method 2	**Unambiguous**	**Semi-ambiguous**
Detailed	1867	635
Streamlined	1850	652

Table 5.5: Number of unambiguous and semi-ambiguous assignments produced via the detailed Protocol and streamlined Protocol methods

Of the 155 instances of ambiguity elimination between the Method 1 and Method 2 assignments identified in the previous analysis, the majority (148) are retained when the new Protocol is employed. The time required to perform the assignment in this case was on the order of one or two days for the 18 kDa BlgB. Thus, while there are differences in assignments between the two sets, these are minimal. The time saved by using this more basic implementation, however, is far more significant.

Therefore, using the less detailed assignment approach yields peak assignments which are sufficiently less ambiguous than those acquired using the standard method to give good structures more quickly and the process of assignment can be reduced to a much simpler scheme with little or no change to the eventual result.

Of course, the quickest method would be to also automate the symmetry peak finding, thus giving an effectively fully automatic macro, but to do this would require that all expected peaks be present and picked in all the relevant spectra. The non-ideal nature of these spectra would lead to incomplete peak lists and therefore any fully-automated protocol would be reliant upon data that were deficient. This may not be such a problem when using the HNH-NOESY and CNH-NOESY spectra, where symmetry-related peaks will appear in the less crowded H^{ali}-H^{N} region of the HCH-NOESY, but would be more significant for assignment of the HCH- and HCH_3-edited NOESY spectra. In addition, the streamlined implementation of Method 2 achieves assignment in a time frame short enough to render further time gains in this area somewhat incidental. The

second implementation of the protocol was therefore selected as the optimal and used in all subsequent assignments.

5.2 The Remaining Complementary Experiments

With a working protocol in place, the remaining pairs of spectra were examined to assess their suitability for inclusion into the protocol.

5.2.1 CCH-NOESY

The Fourier-transformed spectrum, shown in Figure 5.9 is clean and of a good quality, however, the resolution of the ^{13}C dimensions is somewhat limited. This is due to the presence of the ^{13}C-^{13}C coupling constants which evolve very quickly, forcing the use of very short acquisition times in these dimensions (< 5ms, cf. ~20 ms for ^{15}N). This problem is common to the F_2 dimensions of both the HCH- and CCH-NOESY spectra and results in particular F_1F_3 plane containing many more peaks than would be found in corresponding planes in HNH- and CNH-NOESY. Significant overlap is therefore present, making it difficult to discern individual peaks and to work with individual strips when assigning. The problem is exaggerated in the CCH-NOESY spectrum since the transfer of magnetization is not 'orthogonal' (as in the CNH-NOESY). This results in the appearance of intense autocorrelation peaks which cause further crowding of the spectra. Such extensive overlap leads to extreme difficulties in using this spectrum as an assignment aid with large regions being too crowded to allow extraction of all the relevant data.

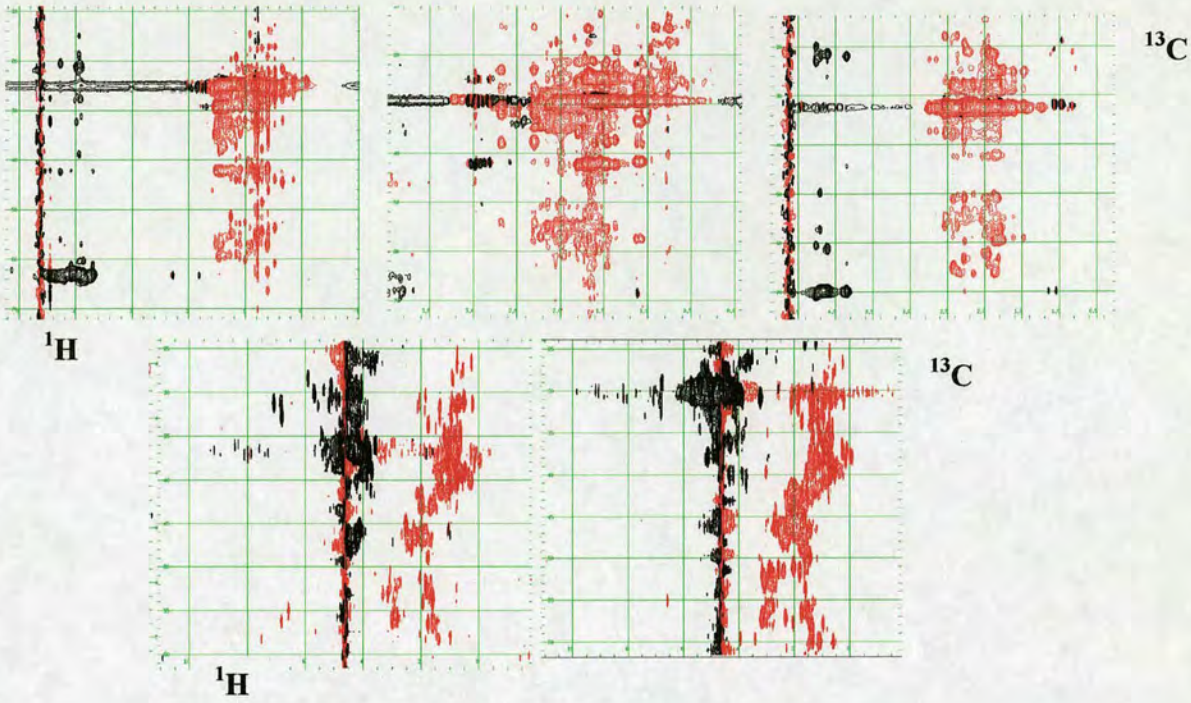


Figure 5.9: F_1F_3 planes (top) and F_2F_3 planes (bottom) of the CCH-NOESY spectrum

In an attempt to improve the resolution of the ^{13}C dimensions, maximum entropy processing was employed. This proved to be non-trivial, however, and, although various parameters were varied in an effort to improve the spectrum, processing by this method continually produced a spectrum in which large regions did not converge and were rendered unusable. Thus it was concluded that the use of maximum entropy processing available to us through AZARA was unsuitable in this case and that the Fourier-transformed spectrum must be used.

The CCH-NOESY spectrum suffers from limited resolution of the indirect dimensions and the extensive overlap of signals. In the HCH-NOESY spectrum, a large difficulty arises in the assignment of the spectrum due to the high degree of degeneracy which exists in the chemical shifts of aliphatic carbons. If this degeneracy is too high, it prevents unambiguous assignment of the NOE peak.

While the use of the CCH-NOESY spectrum is intended to remove this problem, if that spectrum itself is poorly resolved, the problem will only be compounded and the result will be no different, or perhaps even worse, than assignment without the complementary spectrum. The user must begin to operate by selecting only useable regions, giving a patchwork approach to the implementation of this spectrum in the protocol. If a high degree of uncertainty exists in both spectra, the usefulness of the complementary spectrum as an assignment aid is severely compromised. Of course, it is possible to apply the protocol using multiple F_2F_3 assignments but the degree of ambiguity would not be reduced as significantly as might be hoped and would certainly be less than in the case of well resolved spectra such as the HNH-NOESY and CNH-NOESY spectra. This spectrum was therefore excluded from the protocol.

5.2.2 NNH-NOESY

This spectrum was well resolved but presented a different problem in that, on examining the strips of the spectrum and comparing them to the equivalent ^{15}N-edited NOESY strips, it became apparent that the strips of the NNH-NOESY spectrum consistently contained far fewer peaks than expected. A comparison of the number of H^N-H^N and N^H-H^N peaks present in equivalent strips of the HNH-NOESY and NNH-NOESY spectra is given in Table 5.6. In this case, each peak in the complementary spectrum would be generally be expected to correspond to one peak in the standard spectrum (amide ^{1}H's) or, more rarely, two peaks (amine ^{1}H's).

Spectrum	No. peaks
HNH-NOESY	118[a]
NNH-NOESY	22

Table 5.6: Numbers of NH peaks present in 30 equivalent strips of the HNH-NOESY and NNH-NOESY spectra. [a]This is the number of peaks present in the amide/amine ^{1}H region of each strip.

Clearly, the NNH-NOESY spectrum gives a considerably poor result with only ca. 19% of the number of peaks present in the H^N-H^N region of the HNH-NOESY spectrum. Such a large discrepancy indicates a severe lack of signal in this spectrum and suggests that its use in the protocol may be limited. On further examination, the NNH-NOESY spectrum is seen to contain little to no useful information throughout, with the only significant peaks apparently those of intra-residue and sequential NOE's. Some of the clearest examples are shown in fig 5.10. In these examples it can be seen that, while the HNH-NOESY strips contain at least 5 N^H-N^H crosspeaks in each case, the corresponding NNH-NOESY strips show only one or zero crosspeaks.

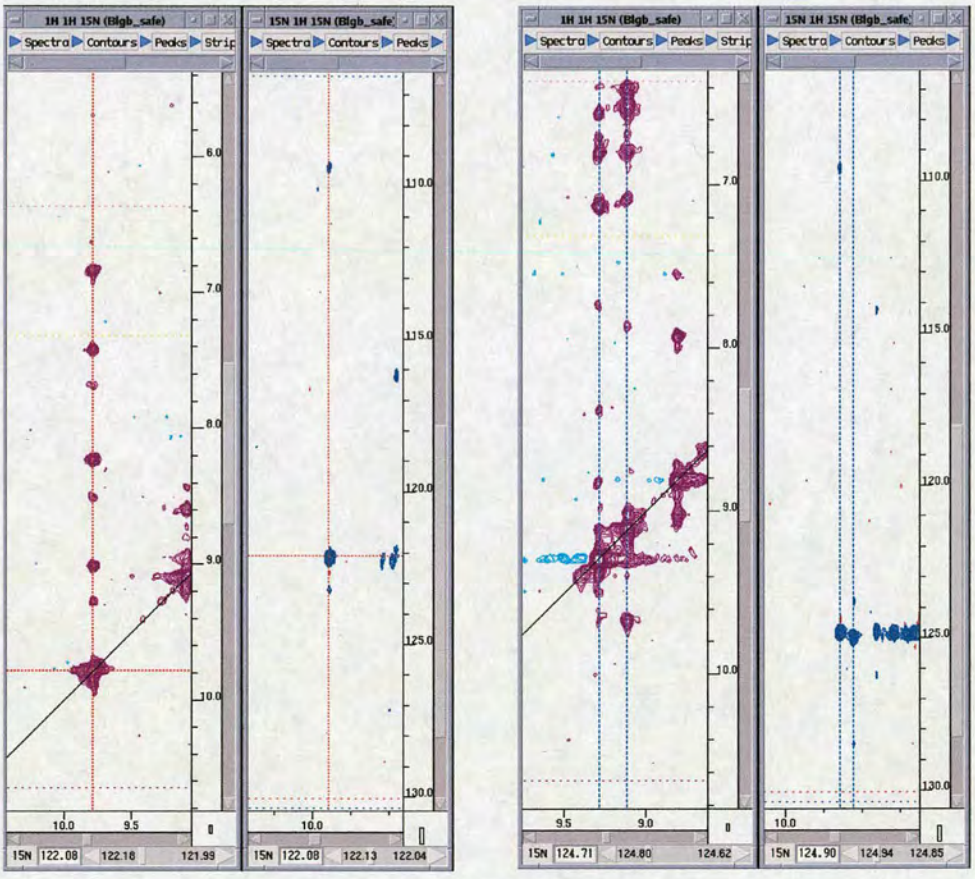


Figure 5.10: HNH-NOESY strips (in purple, showing H^N-H^N region) with equivalent strips of the NNH-NOESY spectrum (shown in blue).

The quantity of signal is clearly insufficient and, thus, filtering assignment possibilities via this spectrum would produce either no unambiguous assignments or, worse, incorrect assignments. Furthermore, the HNH-NOESY spectrum is well resolved and, in general, ^{15}N shifts show a high degree of dispersion, lessening the need for a complementary spectrum to reduce assignment ambiguity. Inclusion of the NNH-NOESY spectrum in the Protocol was therefore deemed ineffective and unnecessary.

5.2.3 CCH_3-NOESY

As with the CCH-NOESY experiment, this experiment also contains two ^{13}C dimensions but, since it only contains signals from methyl groups (and some sidechain CHs), there is far less overlap in the spectrum. The problem of poor resolution in the ^{13}C dimension is effectively removed due to the high digital resolution employed in the second ^{13}C dimension.

The signal density is also favourable, as shown in Table 5.7 with the CCH_3-NOESY strips containing ca. 65 % of the peaks present in the HCH_3-NOESY strips. The HCH_3-NOESY and CCH_3-NOESY were therefore also chosen to be included in the protocol.

Spectrum	No. peaks
HCH_3-NOESY	244[a]
CCH_3-NOESY	160

Table 5.7: Numbers of peaks present in 30 equivalent strips of the HCH_3- and CCH_3-NOESY spectra. [a]This is the number of peaks present in the aliphatic ^1H region of each strip

5.3 Assignment of HCH_3-NOESY

The HCH_3-NOESY spectrum was assigned with the aid of the CCH_3-NOESY spectrum using the streamlined Protocol (section 5.1.3). Again, the HCH-NOESY spectrum was used to look for symmetry related peaks. As with the HNH-NOESY spectrum, assignment was carried out simultaneously via Method 1. The results of assignment via the two methods are given in Table 5.8.

Method	Unambiguous	Semi-ambiguous	Ambiguous
1	852	617	330
2	968	487	344

Table 5.8: Comparison of peak assignments in the HCH_3-NOESY spectrum obtained via Methods 1 and 2. These are the totals prior to removal of erroneous and duplicate assignments.

As with the HNH-NOESY spectrum, assignment via Method 2 generates more unambiguous assignments. Prior to calculating the number of cases of ambiguity elimination and reduction, erroneous assignments were removed by performing a retrospective analysis of the restraint data using the final structures of BlgB given in section 6.4. Duplicate restraints present in the datasets were also removed. The number of ambiguity elimination and reduction events, and their NOE classes were then calculated. The results are shown in Table 5.9.

Event	Sequential	Short	Long
Ambiguity Elimination	9	39	75
Ambiguity Reduction	12	60	65

Table 5.9: Breakdown of ambiguity elimination and reduction events identified in assigning the HCH_3-NOESY spectrum via Method 2 compared to assignment via Method 1.

In total, the ambiguity elimination and reduction events represent ca. 25 % of the CH_3-H^{ali} restraints.

Thus, as was seen for the HNH-NOESY, assignment via Method 2 gives restraints which are not only less ambiguous than their standard method counterparts but, of the less ambiguous restraints, the majority are concentrated in the short- and long-range NOE types. The significance of the ambiguity elimination events is also supported by a QUEEN analysis of the (correct) unambiguous restraints produced by Method 2, shown in Figure 5.11. Again, as was the case with the HNH-NOESY, these events are also seen to be among the most unique pieces of information present in the dataset.

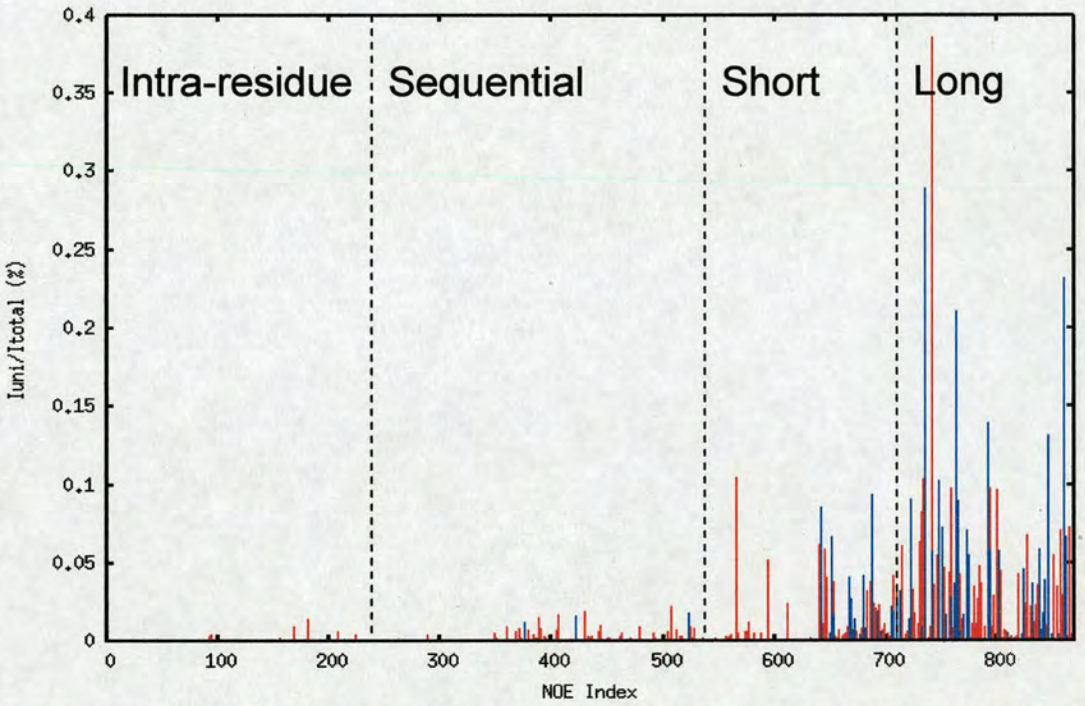


Figure 5.11: Plot of I_{uni}/I_{total} values for unambiguous HCH_3-NOESY restraints obtained via Method 2. Restraints which correspond to ambiguity elimination events are in blue

Furthermore, the number of incorrect assignments obtained via Method 2 is also less than for Method 1. In total, of the 2803 unambiguous HNH- and HCH_3-NOESY restraints obtained via Method 2, 123 (ca. 4%) were found to be incorrect whereas, for Method 1, 221 of the 2572 unambiguous restraints were incorrect (ca. 9%).

5.4 The Protocol thus far

Of the four complementary BlgB spectra which were acquired and tested, two were identified as the most useful spectra for inclusion in the protocol owing to their good resolution and suitable signal density. These were the CNH-NOESY (for use with the HNH-NOESY) and the CCH_3-NOESY (for use with the HCH_3-NOESY).

Of the two implementations of Method 2 which were tested, the second was chosen to be the most useful. This implementation provided almost identical assignments to those obtained via the more detailed first implementation and in much less time.

Assignment of both the HNH- and HCH_3-NOESY spectra via Method 2 resulted in a modest increase (ca. 7% for HNH-NOESY and 14% for HCH_3-NOESY) in the number of unambiguous assignments relative to assignments obtained via Method 1. However, these additional unambiguous assignments were largely comprised of more structurally significant NOEs and the information content of these restraints was also shown to be among the highest in each dataset. This resulted in preliminary structures with better convergence being obtained more quickly than in the same process conducted using the Method 1 restraints. The reduced ambiguity events also formed a significant portion of the total restraint data (30% of the semi-ambiguous assignments obtained via Method 2 were less ambiguous than their Method 1 counterparts) and, again, improved the efficiency of the structure calculation process relative to Method 1.

The BlgB structures obtained thus far (using only the HNH-NOESY restraints), although of a reasonable quality, still require improvement. The HCH_3-NOESY restraints must be included and their effect on the structures assessed. Furthermore, the information which exists in the HCH-NOESY experiment must also be considered and incorporated into the full protocol in some way. This is the subject of the following chapter.

Chapter 6

BlgB

6.1 Background

β-lactoglobulin is a small homodimeric protein (~18 kDa monomer) found in the milk of many animals, but not in humans (Hambling et al., 1992). Its two most common forms (A and B) differ by only two amino acids (in variant A, Ala118 is Val and Gly64 is Asp) and, of these, the low pH NMR structure has been solved for the A variant (Uhrinova et al., 1998, 2000). Under these conditions, the protein undergoes a structural change and exists in a partially unfolded state as a monomer. X-ray structures of both variants have been solved at a range of pH values close to physiological pH (Oliviera et al., 2001; Brownlow et al., 1997; Qin et al., 1998, 1999). The NMR-derived structures of BlgA show that the low pH monomer retains the majority of structural features of the dimer. The properties of variants A and B are markedly more dissimilar than their sequences, however, exhibiting different aggregation properties during milk processing (Hill et al., 1996). An understanding of the origins of this difference in behaviour is therefore of significant interest to the dairy industry. The amino acid sequence of BlgB is given below and the X-ray structure of BlgB is shown in Figure 6.1, consisting of a calyx formed by 8 β-strands (A-H) with a three-turn α-helix running parallel to the A, F, G and H strands and a final β-strand, I, at the C-terminus. Blg exhibits a number of pH-induced transitions known collectively as the Tanford transition (Tanford, 1959), with the principal structural difference being the position of the EF loop which, at low pH, moves to cover the opening of the calyx (Qin et al., 1998).

-3 VPM
1 LIVTQTMKGL DIQKVAGTWY SLAMAASDIS LLDAQSAPLR
41 VYVEELKPTP EGDLEILLQK WENGECAQKK IIAEKTKIPA
81 VFKIDALNEN KVLVLDTDYK KYLLFCMENS AEPEQSLACQ
121 CLVRTPEVDD EALEKFDKAL KALPMHIRLS FNPTQLEEQC
161 HI

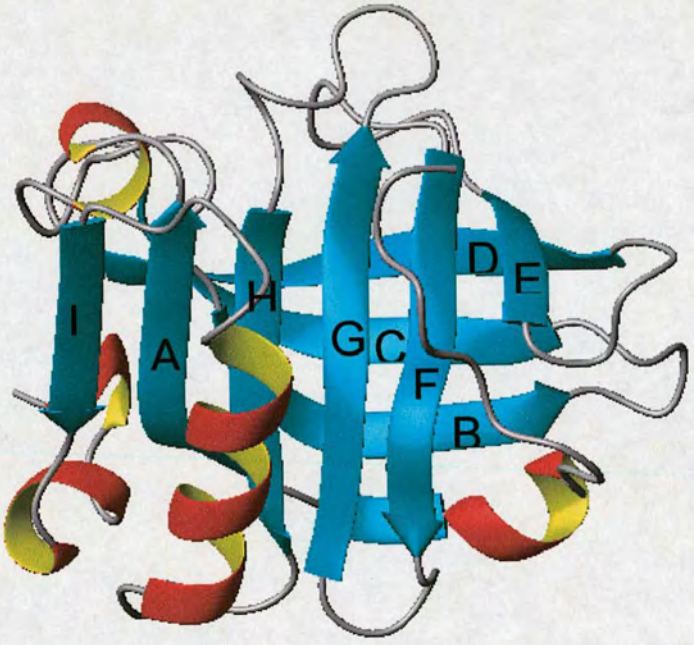


Figure 6.1: Cartoon representation of the X-ray structure of BlgB generated using MolMol

6.2 Experimental

6.2.1 Resonance Assignment Experiments

The resonance assignment experiments detailed in section 3.3 were acquired on the BlgB (monomer) sample on a Bruker Avance 600 MHz spectrometer equipped with a 5mm triple-axis gradient, triple resonance probe. All spectra were recorded at 310 K. The offsets used for ^{1}H and ^{15}N were 4.63 ppm and 118.3 ppm respectively. For ^{13}C, the

offset was 41.50 ppm except for the HCCH-TOCSY (66.0 ppm), (HB)CB(CGCD)HD & (HB)CB(CGCDCE)HE (34.8 ppm) and aromatic HSQC (124.8 ppm) experiments. Spectral folding was employed for all ^{15}N dimensions and for the ^{13}C dimension of the HCCH-TOCSY. The acquisition parameters are summarised in Table 6.1.

Experiment	No. Dims	No. points	Sweep Widths (Hz)
^{15}N-^{1}H-HSQC	2	1024 x 64	8389 x 1458
HBHA(CO)NH	3	1024 x 64 x 96	8389 x 1458 x 6400
CBCA(CO)NH	3	1024 x 64 x 128	8389 x 1458 x 11312
HNCACB	3	1024 x 64 x 128	8389 x 1458 x 11312
H(C)(CO)NH	3	1024 x 70 x 128	8389 x 1458 x 4198
(H)C(CO)NH	3	1024 x 70 x 128	8389 x 1458 x 11312
HCCH-TOCSY	3	1024 x 64 x 128	8389 x 5731 x 4198
(HB)CB(CGCD)HD	2	1024 x 96	8389 x 5634
(HB)CB(CGCDCE)HE	2	1024 x 96	8389 x 5634
Aromatic HSQC	2	1024 x 248	8389 x 8049

Table 6.1: Acquisition parameters for the resonance assignment experiments acquired for BlgB

All spectra were processed using the program AZARA. For spectra which were Fourier transformed, the data were zerofilled once in each dimension. For the H(C)(CO)NH, (H)C(CO)NH and HCCH-TOCSY experiments, maximum entropy processing was used for the indirectly detected dimensions.

6.2.2 NOESY experiments

All NOESY experiments were acquired on a Bruker Avance 800MHz spectrometer equipped with a 5mm triple axis gradient, triple resonance probe. All spectra were recorded at 310 K. ^{1}H and ^{15}N offsets were 4.62 ppm and 118.3 ppm respectively. Prior to the acquisition of the methyl-selective experiments, a 2D constant-time, methyl-selective ^{13}C-^{1}H-HSQC was acquired to obtain suitable offset and sweep width

parameters for appropriate folding of the methyl region. All NOE mix times were 100 ms.

HNH- HCH- & HCH_3-NOESY

The HNH-NOESY experiment was acquired with 1980 points in the directly detected ^1H dimension (t_3) and with 144 and 80 points in the indirect ^1H (t_1) and ^{15}N (t_2) dimensions respectively. Sweep widths (Hz) were 10416 x 1946 x 9599 (t_3 x t_2 x t_1). 8 scans were acquired for each increment giving a total acquisition time of 72 hours and 45 minutes. The processing of this spectrum was carried out using AZARA using maximum entropy in the indirect dimensions to bring the total size of the spectrum to 2048x128x512 ($F_3xF_2xF_1$).

The HCH-NOESY experiment was acquired with 2048 points in the directly detected ^1H dimension (t_3) and with 224 and 64 points in the indirect ^1H (t_1) and ^{13}C (t_2) dimensions respectively. Spectral folding was employed in the ^{13}C dimension with a ^{13}C offset of 66.0 ppm. Sweep widths (Hz) were 10416 x 7646 x 9599 (t_3 x t_2 x t_1). 16 scans were acquired for each increment giving a total acquisition time of 94 hours and 30 minutes. The processing of this spectrum was carried out using AZARA using maximum entropy in the indirect dimensions to bring the total size of the spectrum to 2048x256x512 ($F_3xF_2xF_1$).

The HCH_3-NOESY experiment was acquired with 1024 points in the directly detected ^1H dimension (t_3) and with 196 and 128 points in the indirect ^1H (t_1) and ^{13}C (t_2) dimensions respectively. Spectral folding was employed in the ^{13}C dimension with a ^{13}C offset of 17.3 ppm. Sweep widths (Hz) were 10416 x 4930 x 9599 (t_3 x t_2 x t_1). 8 scans were acquired for each increment giving a total acquisition time of 86 hours. The processing of this spectrum was carried out using AZARA using maximum entropy in the indirect dimensions to bring the total size of the spectrum to 1024x512x512 ($F_3xF_2xF_1$).

CNH- and CCH_3-NOESY

The CNH-NOESY experiment was acquired with 1280 points in the directly detected ^{1}H dimension (t_3) and with 160 and 80 points in the indirect ^{13}C (t_1) and ^{15}N (t_2) dimensions respectively. Sweep widths (Hz) were 10416 x 1946 x 15094 (t_3 x t_2 x t_1). The ^{13}C offset was set to 41.5 ppm; no spectral folding being used in this dimension. 16 scans were acquired for each increment giving a total acquisition time of 88 hours. The processing of this spectrum was carried out using AZARA using maximum entropy in the indirect dimensions to bring the total size of the spectrum to 2048 x 128 x 256 ($F_3xF_2xF_1$).

For the CCH_3-NOESY experiment, acquisition was performed with 2048 points in the directly detected dimension, 128 points in the indirect methyl-^{13}C dimension (t_2) and 160 points in the indirect ^{13}C-aliphatic dimension (t_3). 16 scans were acquired for each increment giving a total acquisition time of 76 hours 30 minutes. Sweep widths (Hz) were 10416 x 4930 x 15094 (t_3 x t_2 x t_1). Spectral folding of the $^{13}CH_3$ dimension was employed with a ^{13}C offset of 17.3 ppm. For the first ^{13}C HSQC the ^{13}C offset was set at 41.5 ppm. The processing of this spectrum was carried out using AZARA using maximum entropy in the indirect dimensions to bring the total size of the spectrum to 2048 x 128 x 512 ($F_3xF_2xF_1$).

6.3 Resonance assignment

6.3.1 Backbone

The majority of the backbone resonance assignment of BlgB was carried out by Patrick Edwards using the CBCA(CO)NH, HNCACB and HBHA(CO)NH experiments. No assignments were made for residues -3 to 3 at the N terminus and, at the C terminus, residues 153-162 were also unassigned. Prior to commencing the sidechain assignment, these spectra were investigated again in order to acquire backbone assignments for these

residues. The assignment program ANSIG (Kraulis, 1989) was used to assign these spectra. After a close inspection, assignments were obtained for residues -1 to 3 and residues 152 to 158. The ^{15}N-^{1}H HSQC spectrum is shown in Figure 6.2

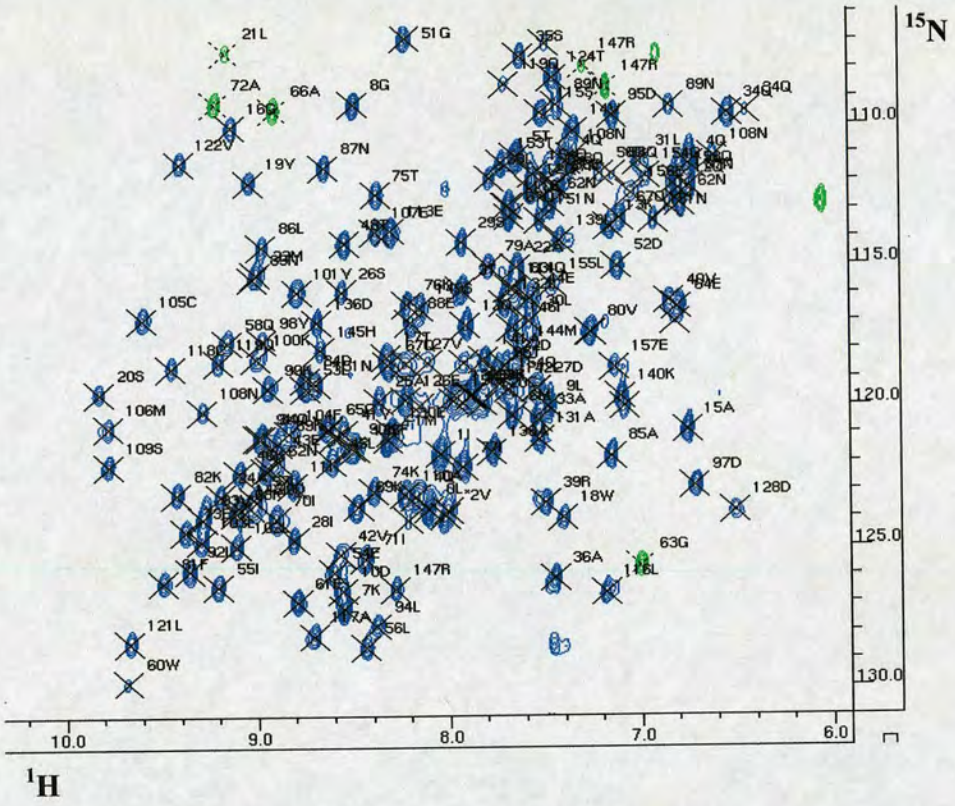


Figure 6.2: ^{15}N-^{1}H-HSQC of BlgB. Aliased peaks appear negative (shown in green)

HNCACB strips illustrating the connectivities established for residues 154 to 158 are shown in Figure 6.3 and the corresponding HBHA(CO)NH strips are shown in Figure 6.4.

Backbone assignments were thus obtained for all but 5 residues (-3V, 159Q-162I). While there remain some unidentified peaks in the ^{1}H-^{15}N-HSQC spectrum, the corresponding strips of the backbone experiments contained too few peaks to allow further connectivities to be established and thus assignment was not possible for these. No experiments were acquired for assignment of the carbonyl groups.

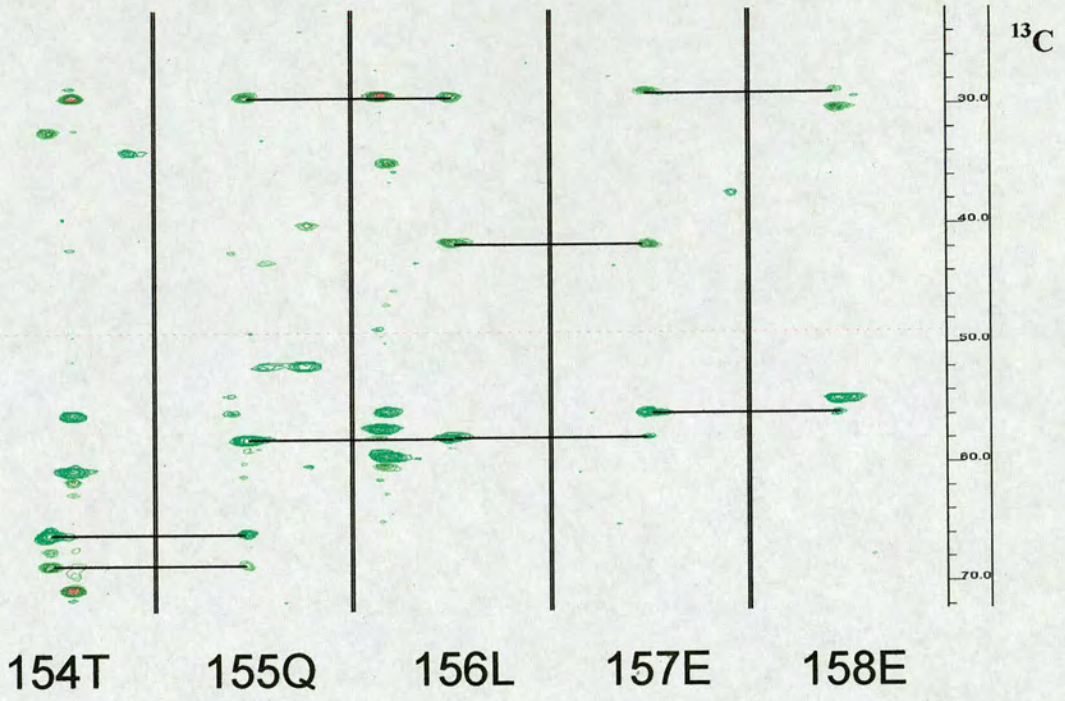


Figure 6.3: Strips of HNCACB spectrum of BlgB

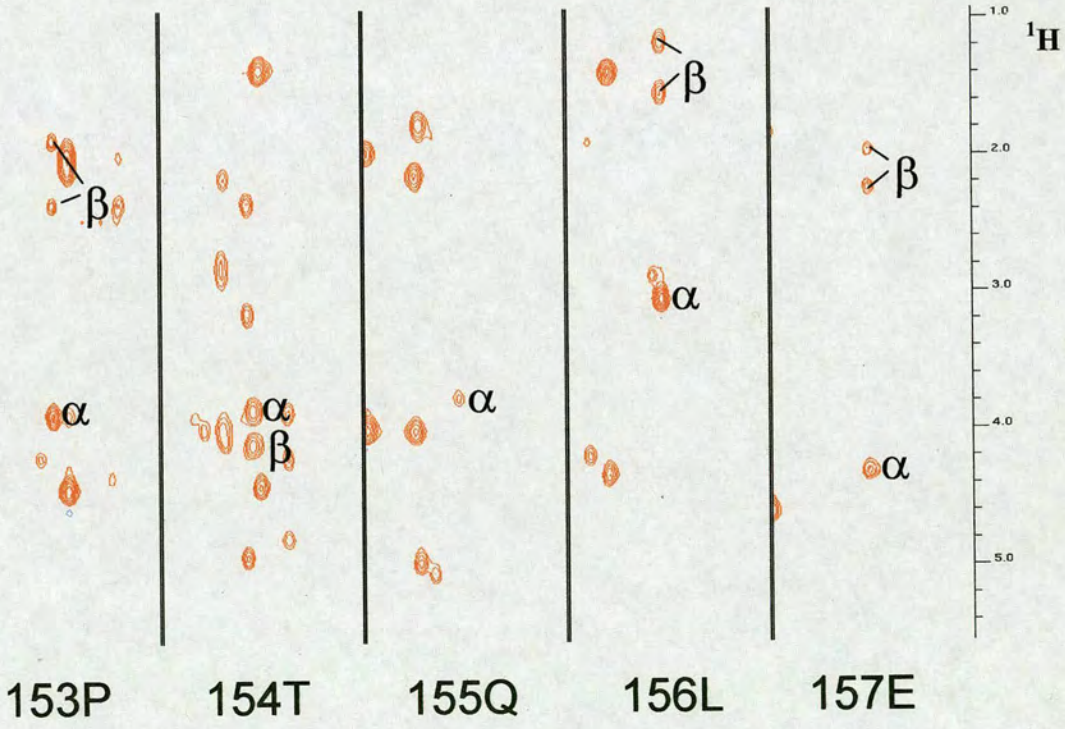


Figure 6.4: Strips of the HBHA(CO)NH spectrum of BlgB

6.3.2 Sidechain Assignment

For the sidechain assignments, the H(C)(CO)NH, (H)C(CO)NH and HCCH-TOCSY experiments were acquired and assigned as described previously. Figures 6.5 a) and b) show representative strips of the first two spectra focusing on the newly assigned residues. Assignment of the H(C)(CO)NH and (H)C(CO)NH experiments yielded the majority of the sidechain resonances but, in some cases, it was not possible to identify a particular resonance due to the signal overlap. In these cases, use of the HCCH-TOCSY enabled identification of the individual resonances. Assignment of the sidechain amine groups was achieved by referring to the ^{15}N-edited NOESY spectrum as described. In this way, full sidechain assignments were obtained for 111 (of 153) non-aromatic residues. Of the remaining, partially assigned residues it was not possible to identify peaks corresponding to the majority of the exchangeable sidechain resonances. Of the Arg residues, none of the N^{ε}, H^{ε} and branching NH_2 groups were assigned, and none of the N^{ζ} and H^{ζ} resonances of Lys residues were identified. OH resonances of Ser and Thr were also unassigned. The assignment of C^{β} of the Trp residues was not possible as only the C^{α} nuclei of these residues were ^{13}C-labelled.

6.3.3 Aromatic assignments

Assignment of the aromatic sidechain resonances was achieved using the (HB)CB(CGCD)HD, (HB)CBCGCDCE)HE and aromatic-HSQC spectra, which are shown in Figure 6.6.

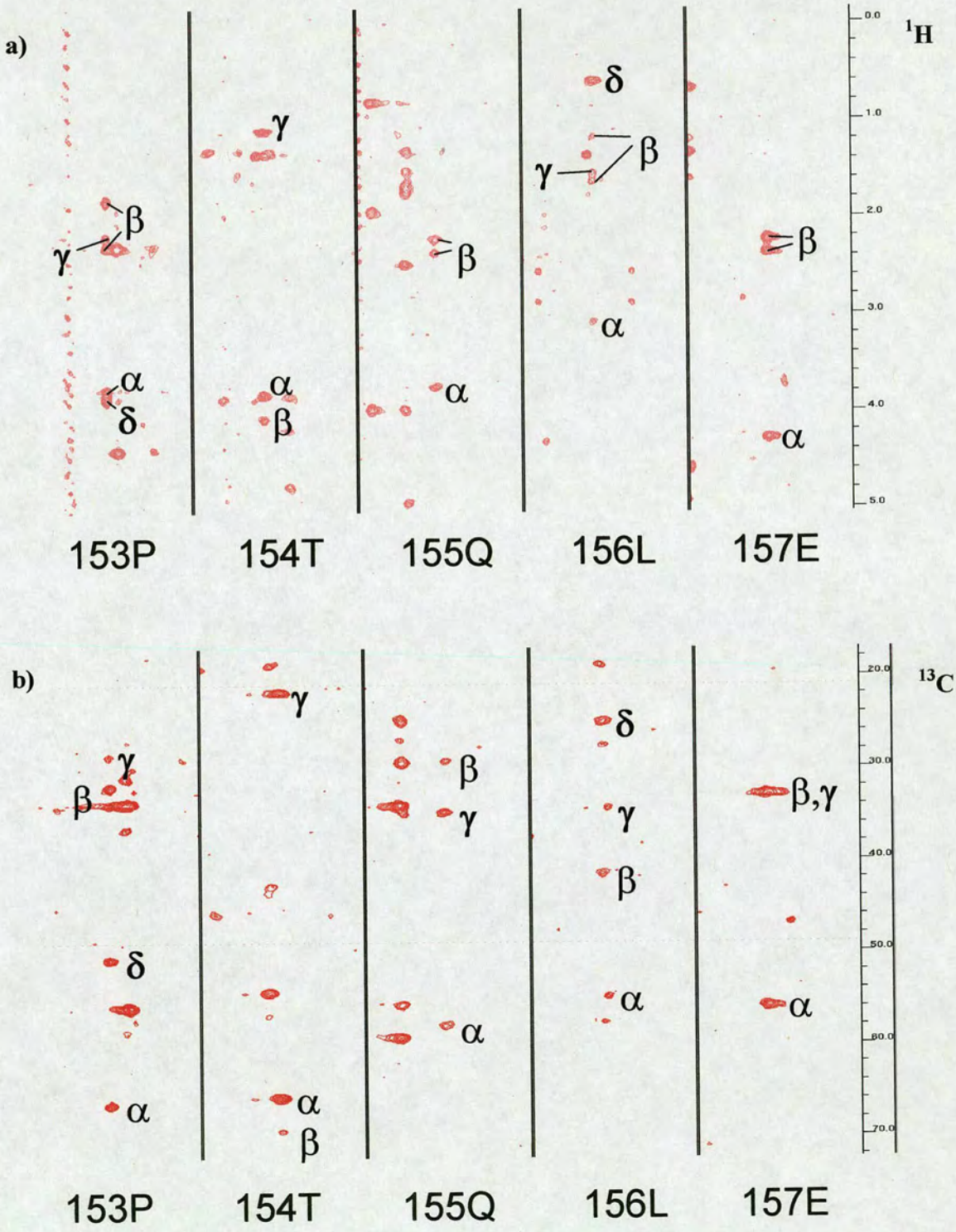


Figure 6.5: Strips of a) the H(C)(CO)NH and b) the (H)C(CO)NH spectra

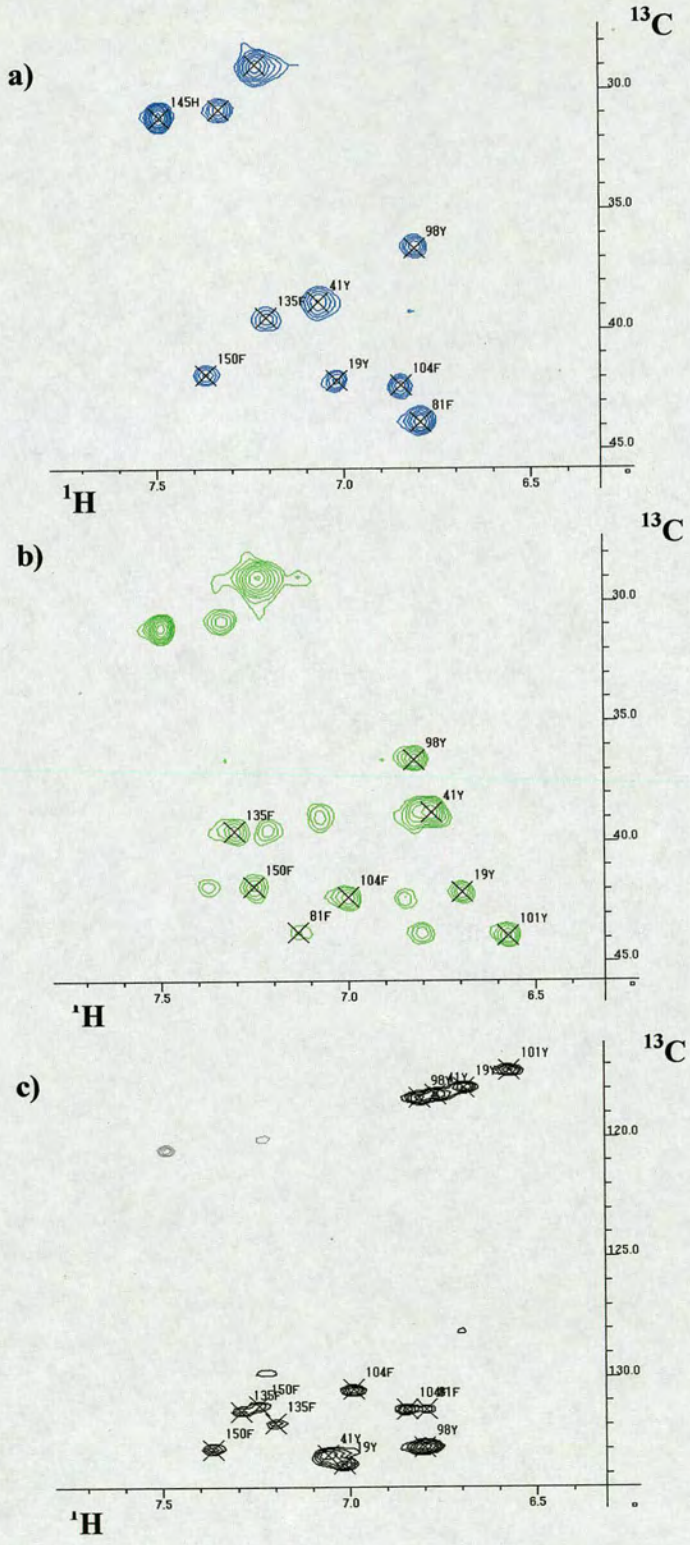


Figure 6.6: a) (HB)CB(CGCD)HD b) (HB)CBCGCDCE)HE and c) Aromatic HSQC spectra

The (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE provided H^{δ} and H^{ε} assignments for all Phe and Tyr residues with the exception of 101Tyr, whose H^{δ} resonance overlapped with that of Phe81. The Aromatic-HSQC experiment was therefore required to determine the chemical shift of this resonance. The labile Tyr OH protons do not appear. Assignment of the four Phe H^{z} protons was not possible due to their signals being too closely overlapped (not shown in figure). Of the two His residues, only one had a C^{β} assignment and, thus, only this H^{δ} was identified. Since only the C^{α} of Trp residues were ^{13}C enriched, the Trp aromatic signals could not be assigned in this way. While other methods could be used to obtain the Trp ^{1}H sidechain resonances, it was decided that this was unnecessary as the BlgB sequence contains only two Trp residues.

6.3.4 Assignment Summary

With the exception of $^{13}C'$ resonances and the labile OH resonances complete assignment was achieved for 111 of 165 residues. If the labile Lys and Arg sidechain resonances are also excluded, the total number of completely assigned residues is increased to 129. In total, the backbone assignment was 95% completed and the sidechain assignment 91% complete. This was considered to be near-complete assignment of the protein and, thus, assignment of the NOESY spectrum could commence.

The resulting C^{α}, H^{α} and C^{β} chemical shifts were used to generate a consensus chemical shift index (CSI) to predict the secondary structure elements of the protein, the results of which are illustrated in Figure 6.7.

The secondary structure features predicted by the CSI (Wishart & Sykes, 1994) correspond to the X-ray structure of BlgB with nine β-strands and a single α-helix. This information was used to generate 110 dihedral angle restraints corresponding to those angles most favoured by β-strands and α-helices for use in the structure calculation.

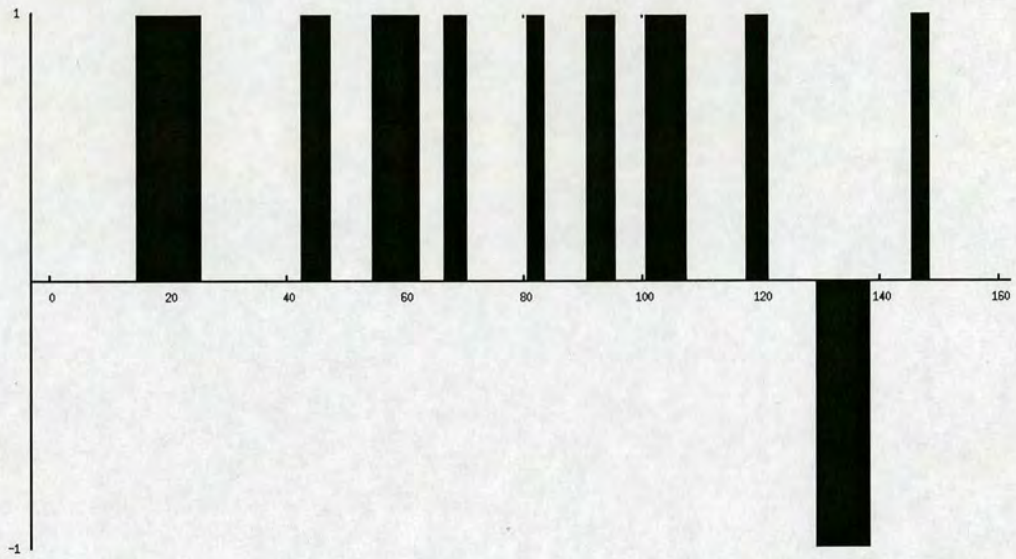


Figure 6.7: Consensus CSI for BlgB. Regions where >3 consecutive residues display an index of 1 indicate β-sheet and -1, α-helix.

6.4 NOESY Assignment and Structure Calculation

The HNH-, CNH-, HCH_3-, CCH_3- and HCH-NOESY spectra were loaded into an Analysis project along with the chemical shift table obtained from the resonance assignment. The assignment of the HNH-NOESY and HCH_3-NOESY spectra was performed via the streamlined implementation of Method 2 as described previously resulting in 2818 unambiguous peaks for use in the first round of structure calculation. The peak assignments were converted into restraints using the 'Make Distance Restraints' function of analysis and exported as ARIA restraint tables. These restraints were used to calculate 100 structures. Following the checking and treatment of violations, the total number of unique, unambiguous HNH- and HCH_3-NOESY peaks was 2510.

The family of structures thus obtained, shown in Figure 6.8a, had somewhat poor backbone rmsds (for residues 3-155) of only 2.1 Å but this is largely due to the less well defined loops. Considering only α-helix and β-sheet secondary structure elements gives a much better rmsd of 1.7 Å although, at this stage, the α-helix was very poorly defined.

The lowest energy structures were then used to filter the semi-ambiguous restraints. In total, 1226 restraints were submitted to the process. In this filtering step, restraints contribution less than 1% of the intensity of the peak were discarded. Prior to the next round of calculations, the violations identified by the checking process were examined. Only one further filtering round was employed, again with restraints contributing less than 1% of the peak intensity being discarded – additional rounds showing no further improvement in the structure. The addition of these restraints gave a much improved structure ensemble (backbone rmsd 1.44 Å) as shown in Figure 6.8b and increased the total number of converged structures to 25.

Thus, after three rounds of calculations, the final set of restraints were added – the completely ambiguous restraints. As before, these restraints were filtered using the structure ensemble of the previous calculation. After two further rounds of calculations, filtering of the ambiguous restraints did not improve the structure of subsequent rounds and thus, this was considered the conclusion of the calculation with respect to the HNH- and HCH_3-NOESY spectra. The resulting ensemble is shown in Figure 6.8c with the energy profile and structural statistics for this ensemble given in Figure 6.9 and Tables 6.2, 6.3 and 6.4. 30 structures were slected for the analysis on the basis of the jump in the rmsd profile in Figure 6.9.

The structures thus obtained are of good quality; exhibiting low rmsds (0.96 Å, backbone) and the small number of NOE violations indicate that the structures are consistent with the experimental data. However, it is still possible to improve the quality of the structures since a good deal of NOE information remains unused in the HCH-

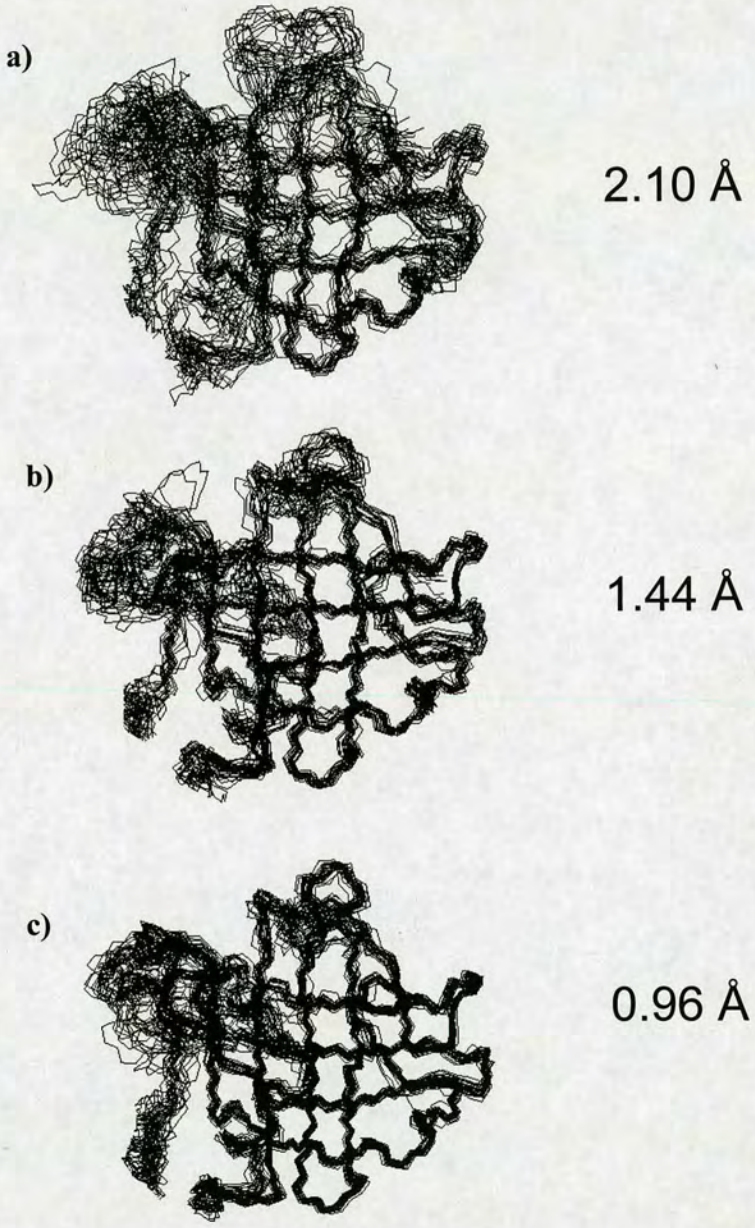


Figure 6.8: Summary of the structure calculations for BlgB performed using the HNH- and HCH_3-NOESY restraint data. Showing an overlay of the lowest energy structures obtained using, a), unambiguous (20/100), b), unambiguous and semi-ambiguous (25/100), and, c), unambiguous, semi-ambiguous and completely ambiguous restraints (30/100), c). Backbone rmsd's are given in each case for residues 3-155. Residues -3 to -1 and 156-161 have been omitted.

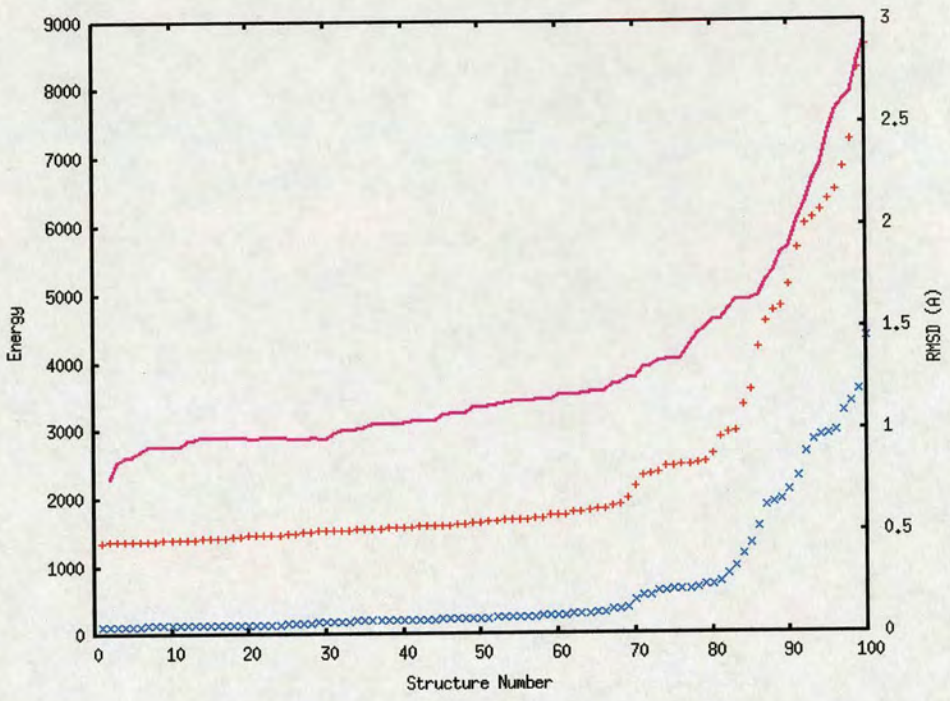


Figure 6.9: Plot of total potential energy (red), NOE energy (blue) and backbone rmsd (for residues 3-155, magenta) for the structures calculated using all HNH- and HCH_3-NOESY restraints.

NOE Type	No. Restraints[a]
Intra-residue	1046
Sequential	1301
Short ($2 \leq i\text{-}j \leq 4$)	418
Long ($i\text{-}j \geq 4$)	909
Ambiguous	822

Atoms[b]	Rmsd[c]
Backbone	0.96 Å (0.75)
Heavy	1.32 Å (1.10)

Violations	
NOE (>0.3 Å)	5
Torsion Angle (>1.5 Hz)	3

Tables 6.2, 6.3 & 6.4: NOE and structure statistics for the ensemble of structures (30 of 100) calculated using all HNH- and HCH_3-NOESY restraints.[a] These are the number of unambiguous and semi-ambiguous restraints [b] Overlaid on residues 3-155 [c]Numbers in parenthesis are the rmsds for overlays on the β-sheets and α-helix only.

NOESY spectrum. Only restraints derived from the HNH- and HCH_3-NOESY spectra have been used thus far. The HCH-NOESY spectrum, which has also been acquired but used only indirectly (to check for symmetry related peaks) will therefore contain much more information that may be used to further refine these structures. Obtaining this extra information is made much simpler at this stage by employing the ensemble of structures as a template for filtering the HCH-NOESY restraints.

The regions of the HCH-NOESY which were selected were those which contained restraint information not present in the HNH- or HCH_3-NOESY spectra i.e. H^{α} to H^{aro} and H^{ali}; H^{aro} to H^{ali}; H^{ali} to H^{aro} and H^{ali}. Regions of the spectrum close to the diagonal were excluded due to the intense signal in this area which would increase the likelihood of picking erroneous peaks. H^{α}-H^{α} and H^{aro}-H^{aro} correlations were therefore not obtained. Regions representing NOEs to H^{N} were not selected since these had already been selected in the HNH-NOESY. The selected regions are represented diagrammatically in Figure 6.10.

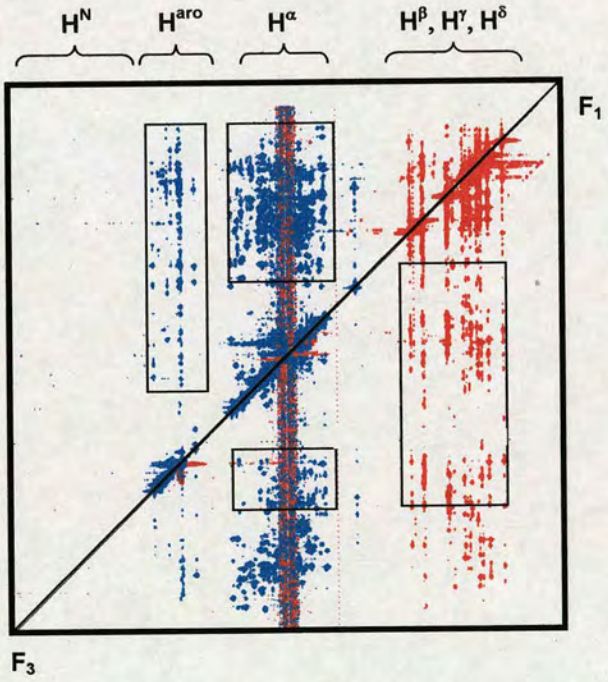


Figure 6.10: Diagrammatic representation of the four regions (in terms of F_1 and F_3) of the ^{13}C-edited NOESY spectrum selected for peak picking.

These regions were submitted to the Region Peak Find feature of the Analysis program to automatically pick the peaks therein, considerably reducing the time spent peak picking. By this method, the user selects a region of the 3D spectrum, specifying the maximum and minimum ppm ranges for each dimension. The program then proceeds to find the maxima (or minima) within that region. To reduce the likelihood that artefacts or noise were selected each plane examined to ensure it was free of any undesirable features. In this way the four regions were examined and picked, yielding a total of 3447 peaks.

This was by no means the total potential information but represents a large portion of the available data and, importantly, required no laborious manual peak picking. It should be noted here that while this appears to be a significant increase in the amount of restraint data, some of it will be degenerate due not only to duplication of restraints within the HCH-NOESY spectrum but also to the duplication of peaks already present and assigned in the HCH_3-NOESY spectrum. Once the peaks had been picked, the planes were manually re-examined to confirm that only 'true' peaks had been picked. *The peaks were not assigned in any dimension.* These ambiguously assigned peaks were then exported as restraints using the 'Shift Match Distance Constraints' function of Analysis and filtered using the structure ensemble obtained previously. In total, 2407 unique restraints were used. Progressive filtering of these restraints in a further three rounds of calculations gave the final structure ensemble. A further structure refinement step was then performed by recalculating this final ensemble in water.

The additional HCH-NOESY restraints result in an increased number of converged structures (60) and improved rmsds as shown in figures 6.11 and 6.12 and in Table 6.4. Thus, while the structure obtained using the HNH- and HCH_3-NOESY restraints was of good quality, a significant improvement (33% reduction in backbone rmsd) is obtained on introducing the restraints from selected regions of the HCH-NOESY. This improvement is achieved quickly and easily via the use of the existing structure ensemble.

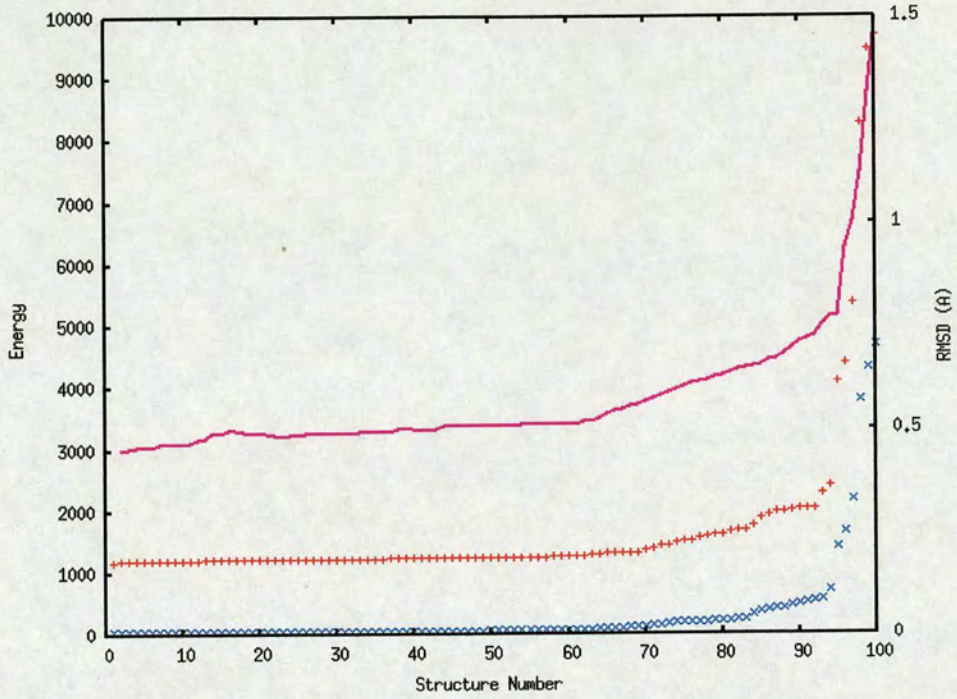
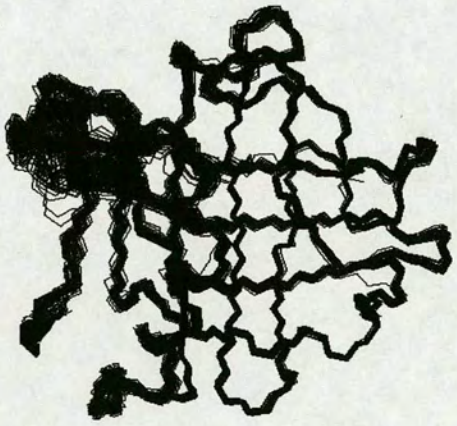


Figure 6.11: Plot of total potential energy (red), NOE energy (blue) and backbone rmsd (for residues 3-155, magenta) for the structures calculated using all HNH- HCH_3- and HCH-NOESY restraints.



Atoms	Rmsd (Å)[a]
Backbone	0.64 Å (0.44)
Heavy	1.04 Å (0.86)

Figure 6.12 & Table 6.4: Overlay (on residues 3-155) and rmsds of the final structure ensemble of BlgB (60/100). [a] Numbers in parenthesis are the rmsds for overlays on the β-sheets and α-helix only.

The full and final Protocol therefore comprises assignment of HNH- and HCH_3-NOESY spectra via the streamlined implementation of Method 2 and subsequent structure calculation followed by the addition of filtered, completely ambiguous HCH-NOESY data. This full Protocol is assessed again, on a second protein, in the next chapter.

It should be noted that BlgB is principally composed of β-sheets. For α-helical proteins, the H^{α}-H^{α} crosspeaks which appear near the diagonal (and, in this case, were not picked) would likely need to be considered. This would require spending more time manually picking the peaks in this region.

6.4.1 PROCHECK & WHATIF

Structure quality checks performed with PROCHECK and WHATIF are given in Tables 6.5-6.7. These show that the structures are of good quality with acceptable values for all the quality tests which were performed. Thus, these analyses indicate that the BlgB structure does not contain unusual (and, therefore, potentially erroneous) elements.

Region	No. residues (%)[a]
Most favourable	63.3 (65.1)
Additionally allowed	28.7 (27.5)
Generously allowed	5.3 (4.7)
Disallowed	2.7 (2.7)

Table 6.5: Ramachandran plot statistics for BlgB calculated using PROCHECK. [a] Numbers in parenthesis are for the BlgA structure calculated by Uhrinova et al (Uhrinova et al., 2000).

Parameter	RMS Z-score
Bond lengths	1.10
Bond angles	0.98
Improper dihedral distributions	1.082
Sidechain planarity	0.97
Omega angles	0.92

Parameter	Z-Score
1^{st} generation packing	-0.745
2^{nd} generation packing	-2.88
χ^1/χ^2 rotamer normality	-2.10

Tables 6.6 and 6.7: RMS Z-scores and Z-scores for WHATIF structure quality checks.

6.5 Structure Analysis

6.5.1 Comparison with X-ray Structure

The overall agreement between the X-ray structure of BlgB (Oliviera et al., 2001) and the closest to mean NMR structure obtained from the ensemble given here is good, showing an average backbone rmsd of 1.2 Å between the two structures when overlaid on residues 3-155. All of the elements of secondary structure present in the X-ray structure are retained in the low pH solution structure. As with the low pH structure of BlgA, the E-F loop is in the 'closed' conformation, occluding the central cavity of the calyx.

Backbone rmsds were calculated for individual regions of the structures (whilst overlaid on the backbone atoms of residues 3-155). The β-strands show the strongest areas of agreement with an average rmsd of 0.9 Å, strand G having the lowest (0.72 Å) and the largest difference being seen in the backbone positions of strand I (1.7 Å). Of the loops and turns, the rmsds are somewhat larger, with an average of 1.3 Å though this may be expected due to the more flexible nature of these regions.

The largest disagreement is found in the orientation of the α-helix and the accompanying β-strand, I. These regions show, on average, a difference of 1.8 Å. Comparison of the backbone positions in the α-helix and strand I alone show a much better agreement of 1.1 Å. The difference is therefore due to the position the α-helix adopts in relation to the calyx. In the NMR structure, the C-terminal end of the α-helix is positioned somewhat closer to the calyx and appears to be angled more toward the N-terminus of the protein as illustrated in Figure 6.13. It can be seen that the α-helix clearly deviates in its position and direction between the X-ray (dimer) and NMR (monomer) structures. In the Blg dimer, strand I forms part of the dimer interface and, thus, this difference could be due to a slight conformational change on undergoing the dimer to monomer transition. Similar differences exist in the X-ray and solution structures of BlgA.

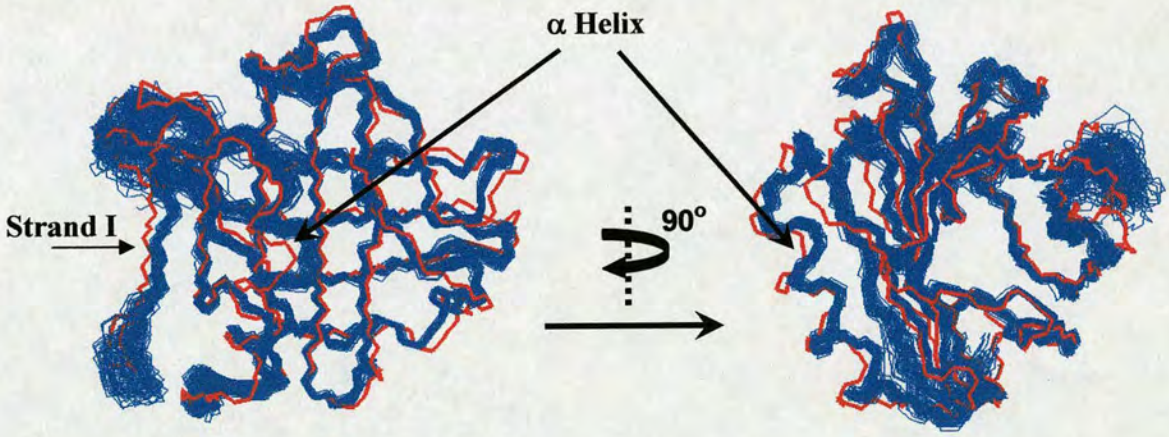


Figure 6.13: Overlay of the NMR-derived structure ensemble (blue) with the X-ray structure (red) of BlgB. Differences in the position of β-strand I and the α-helix are indicated.

6.5.2 Comparison of BlgA and BlgB

As shown in Table 6.5, the NMR-derived BlgA and BlgB structures show similar distribution of Ramachandran plots, although the structure of BlgB shows a slightly improved refinement in terms of the ensemble rmsds – backbone rmsds for the BlgA ensemble of Uhrinova were 0.84 Å (cf. 0.64 Å for BlgB).

The structures of BlgA and BlgB determined by X-ray show little difference (Qin et al., 1999, Oliviera et al., 2001). Of the two differences in the amino acid sequences of the proteins, the 118Val→Ala mutation occurs in the β-strand, H, and has been shown to have little or no effect upon the surface of the protein. It seems likely, therefore, that the differences observed in the physical properties of the two proteins is due to the 64Asp→Gly mutation which occurs in the loop joining the C and D strands. Unfortunately, the flexible nature of this region makes it difficult to observe any particular structural difference this may give rise to. On comparing the NMR-derived structures of variants A and B, the two structures show close agreement with one another (backbone rmsd of 1.3 Å for closest to mean structures) but, unfortunately, the potential area of difference is too poorly defined to allow any useful comparisons to be made.

Chapter 7

C4BP~1,2

7.1 Background

C4b-binding protein (C4BP) is a plasma-based regulator of the complement system, a major factor in innate immune response to foreign pathogens. C4BP regulates both the classical and alternative pathways of complement activation. In its most commonly occurring isoform, C4BP is a 570 kDa protein composed of complement control protein (CCP) modules, arranged in eight chains, linked via their C-termini at a central core region (Dahlbaeck et al., 1983). Of the eight chains, seven are identical α-chains, eight CCP modules in length and the eighth is a shorter β-chain of three CCP modules. These CCP modules are short consensus sequence repeats, ca. 60 aa in length, which form the basis of the majority of complement regulator proteins. Investigations into the C4b binding activity identified the first three N-terminal modules (CCP1-3) of the α-chains as those necessary for binding (Blom et al., 2001). The solution structure of a construct of modules one and two (C4BP~1,2) has recently been solved via NMR (Jenkins et al., 2005). The amino acid sequence of C4BP~1,2 is shown below:

0 M
1 NCGPPPTLSF AAPMDITLTE TRFKTGTTLK YTCLPGYVRS
41 HSTQTLTCNS DGEWVYNTFC IYKRCRHPGE LRNGQVEIKT
81 DLSFGSQIEF SCSEGFFLIG STTSRCEVQD RGVGWSHPLP
121 QCEILEHHHH HH

Naturally, it is necessary to further analyse the Protocol to determine its usefulness in a variety of differing situations. In order to do this, the structure of C4BP~1,2 was solved again. C4BP~1,2 differs from Blgb in a number of ways which make it a good candidate for evaluating the efficacy and applicability of the Protocol. Firstly, C4BP~1,2, as with all CCP modules, has many fewer elements of secondary structure, generally only exhibiting small regions of β-strands. Thus, recalculating the structure of C4BP~1,2 via the Protocol will give some assessment of its performance in solving the structures of structurally varied proteins. Secondly, the number of methyl containing residues (38/133) is also fewer for C4BP~1,2, (65/165 for BlgB) providing a measure of the protocol's ability to perform with reduced data from the methyl-selective experiments. Furthermore, from binding studies performed with M proteins (Jenkins et al., 2005), the residues which were identified as participating in the binding lie on different faces of the two modules, CCP1 and CCP2, leading to the suggestion that there is some re-orientation of the modules upon binding. Further investigation of the solution structure of C4BP~1,2 is therefore of interest.

7.2 Experimental

All experiments were acquired on a Bruker Avance 800 MHz spectrometer, equipped with triple axis gradient5mm triple resonance coil, at 310K. ^{1}H and ^{15}N offsets were 4.63 ppm and 119.3 ppm respectively. All NOE mix times were 100 ms.

The acquisition of the complementary spectra was carried out some months later than the initial ^{13}C-edited and ^{15}N-edited NOESY spectra, during which time the sample had been stored at 4°C. It was therefore necessary to ensure that the sample quality had not deteriorated significantly during this time and that the protein fold remained unchanged to be confident of the validity of using both sets of data together. A comparison of two ^{15}N-^{1}H-HSQCs, one acquired prior to the storage of the sample and the other acquired subsequently is shown in Figure 7.1.

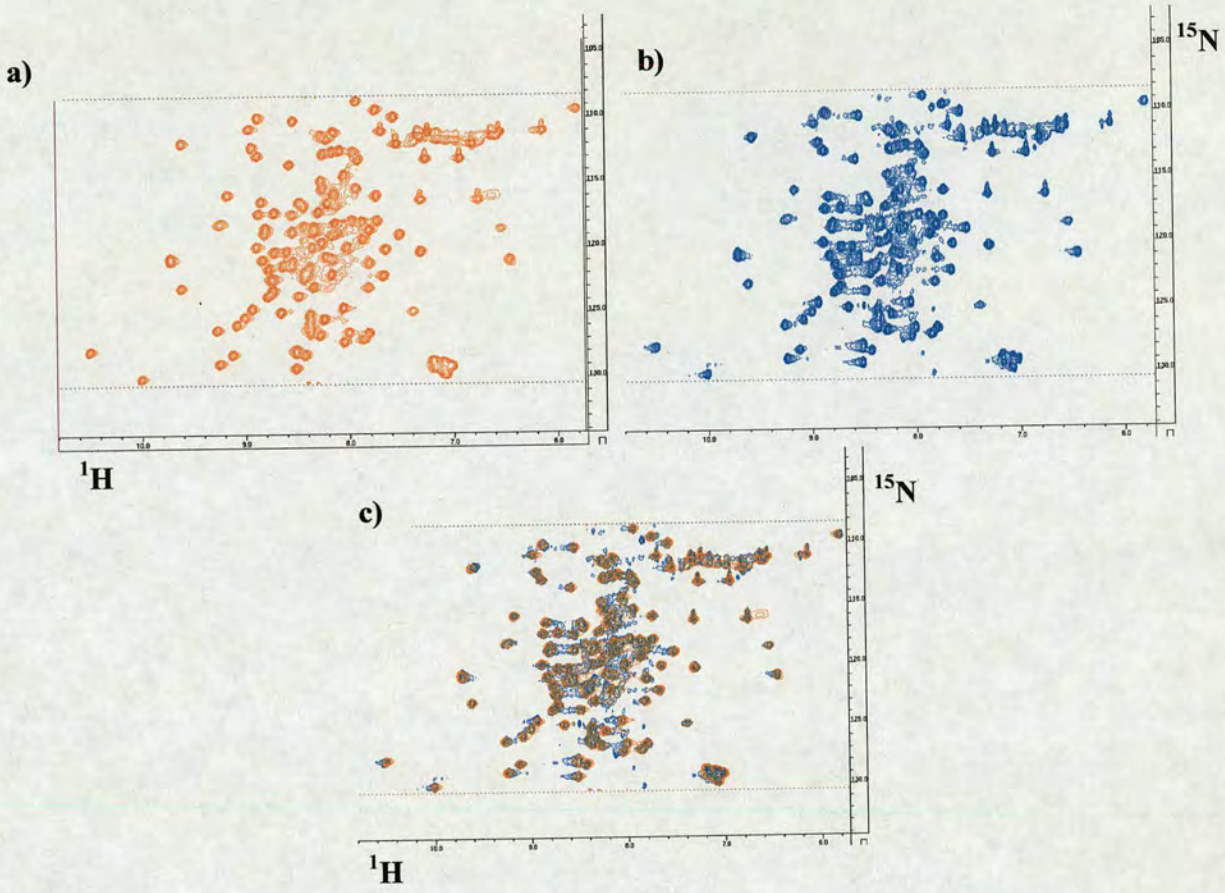


Figure 7.1: ^{15}N-^{1}H-HSQC spectra of C4BP~1,2 acquired before, a) and after storage, b). An overlay of the two spectra is also shown, c). The spectra were plotted with the signal threshold just above the noise level.

From this comparison, it would appear that some degradation of the sample has occurred. The original spectrum is both cleaner than the more recently acquired spectrum and shows peaks which are more intense and well-defined. The increase in the amount of noise around the centre of the spectrum indicates an increase in the proportion of unfolded protein. Comparing the intensities of the NHD peaks (which, in 10% D_2O, account for ca. 20% of the sample) and the spurious peaks it was concluded that the sample was 90% pure. An overlay of both spectra gives further confirmation that the state of the sample is largely unaltered with peak positions showing good agreement. It was therefore decided to continue with the collection of further spectra.

7.2.1 HNH-, HCH- and HCH_3-NOESY spectra

The HNH- and HCH-NOESY spectra used were those acquired previously by Huw Jenkins. Spectral folding was used in both ^{15}N and ^{13}C dimensions, with a ^{13}C offset of 52.9 ppm. These spectra were processed in AZARA (Boucher, 1993) with maximum entropy processing employed in the indirect dimensions. The resulting spectral sizes were thus 2048 x 256 x 512 (^{15}N) and 2048 x 128 x 512 (^{13}C).

The HCH_3-NOESY was acquired with 1782 points in the directly detected dimension, 228 points in the indirect ^1H dimension (t_2) and 112 points in the indirect ^{13}C dimension (t_1). Sweep widths (Hz) were 11160 x 3581 x 9603 (t_3 x t_2 x t_1). 8 scans were acquired per increment giving total acquisition time of 85 hours 45 minutes. Spectral folding of the ^{13}C dimension was achieved with a ^{13}C offset of 20.3 ppm. Following AZARA processing with zerofilling of the t_3 data and maximum entropy processing of the indirect dimensions, the total size of the spectrum was 2048x256x512 ($F_3xF_2xF_1$).

7.2.2 Complementary spectra

Prior to the acquisition of the methyl-selective experiments, a 2D constant-time, methyl-selective ^{13}C-^1H-HSQC was acquired to obtain suitable offset and sweepwidth parameters for appropriate folding of the methyl region.

The CNH-NOESY spectrum was acquired with 2048 points in the directly detected ^1H dimension (t_3) and with 144 and 80 points in the indirect ^{13}C (t_1) and $^{15}N(t_2)$ dimensions respectively. Sweep widths (Hz) were 11160 x 1800 x 15094 (t_3 x t_2 x t_1). 16 scans were acquired for each increment giving a total acquisition time of 88 hours. The ^{13}C offset was set to 44.8 ppm with no folding of the ^{13}C dimension being used The processing of this spectrum was carried out using AZARA using maximum entropy in the indirect dimensions to bring the total size of the spectrum to 2048x256x512 ($F_1xF_2xF_3$).

For the CCH_3-NOESY spectrum, acquisition was performed with 1782 points in the directly detected dimension, 96 points in the indirect methyl-^{13}C dimension (t_2) and 144 points in the indirect ^{13}C-aliphatic dimension (t_3). Sweep widths (Hz) were 11160 x 3581 x 15094 (t_3 x t_2 x t_1). 16 scans were acquired per increment giving a total acquisition time of 90 hours 30 minutes. Spectral folding of the $^{13}CH_3$ dimension was employed with a ^{13}C offset of 20.3 ppm. For the first ^{13}C HSQC the ^{13}C offset was set at 41.5 ppm. Again, AZARA processing was carried on the data with zerofilling of the t_3 data and maximum entropy processing employed in the indirect dimensions. The total spectrum size was 2048x256x512 ($F_3xF_2xF_1$).

7.3 Assignment of NOESY spectra

The spectra were imported into an Analysis project along with the C4BP~1,2 chemical shift assignments obtained from the BMRB. As before, the peaks in each spectrum were picked using Analysis' parabolic peak fitting procedure and the dimensions of the bonded nuclei were assigned. These peaks were then used as input for the refined, semi-automated assignment macro as detailed previously. During this process, the chemical shifts of four sidechain resonances were reassigned. The values for $22N^{\varepsilon}$, $22H^{\varepsilon}$, $105N^{\varepsilon}$ and $105H^{\varepsilon}$ were deemed to be incorrect after examination of the assignment possibilities for their ^{15}N-edited NOESY peaks showed little to no intra-residue or sequential assignments. Examination of the assignment possibilities returned by the protocol gave strong indication that these resonances should be swapped – the $22N^{\varepsilon}$,H^{ε} strip containing numerous assignments for atoms of residues 104, 105 and 106 and vice versa.

In a few cases (16IleH^N and 89GluH^N; 31TyrH^N and 106CysH^N; 42SerH^N and 101SerH^N) due to the close proximity of both their $^{15}N^H$ and $^1H^N$ chemical shifts, the HNH-NOESY strips belonging to these atoms were left unassigned in all three dimensions.

Of the 2497 peaks picked in the HNH-NOESY spectrum, 2111 were assigned, of which 1308 were unambiguous assignments and 803 were semi-ambiguous. For the HCH_3-NOESY, a total of 1428 peaks were selected, 498 of which were assigned unambiguously and 535 semi-ambiguously. The Analysis program was used to convert these assignments into distance restraints which were exported as restraint tables for use as input in the structure calculations performed using CNS. This process was completed within four days.

7.4 Structure Calculation

Following the Protocol, only the unambiguous restraints from the HNH- and HCH_3-NOESY spectra were used as input in the first calculation, along with the four disulphide bridges defined by homology, giving a total of 1806 NOE restraints.

The restraint violations resulting from the initial calculations were analysed and the appropriate peaks were re-examined and, where appropriate, reassigned. Duplicate restraints were also removed from the dataset. This resulted in 1386 unique, unambiguous restraints which were then used to calculate 100 structures, the results of which are given in Figure 7.2. The overall fold of each module agrees with those of the original calculation and of CCP modules in general although, clearly, the resolution requires improvement with quite poor backbone rmsd values for both modules. The intermodular orientation is, naturally, very poor due to the small number of intermodular NOEs present at this stage.

The resolution of module 1 is notably poorer than that of module 2 due to the fact that this module does not converge on a single conformation. These conformations are energetically indistinguishable, however, and it must be concluded that further NOE restraint information is required to determine which is the correct fold. With insufficient information contained within the unambiguous restraints to provide a convergent

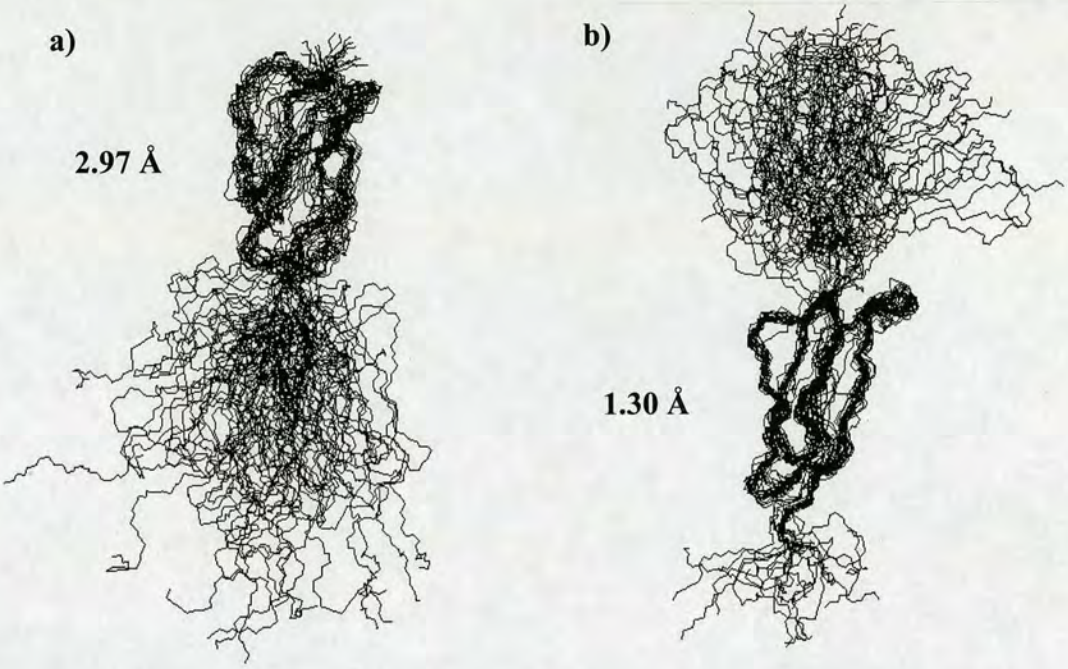


Figure 7.2: Lowest energy (15/100) structures resulting from calculations using unambiguous restraints from both HNH- and HCH_3-NOESY. a) Structures overlaid on module 1 (residues 2-60) and b) overlaid on module 2 (residues 65-122). The numbers indicate backbone rmsds.

structure, the semi-ambiguous restraints were included in the next round of calculations without first filtering them, as was the case with BlgB.

Following this round of calculations, as with all previous rounds, the restraints reported as violations were examined and either reassigned or made ambiguous. Inclusion of the semi-ambiguous restraints gave structures with much improved convergence as shown in Figure 7.3. The inclusion of the semi-ambiguous restraints further improves the rmsd of both modules and all of the lowest energy structures now converge on a single (correct) fold for both modules.

The definition of the intermodular region remains somewhat poor but, again, shows a significant improvement over the previous structure calculation rounds. With a convergent ensemble of structures of sufficient quality, the ARIA protocol could now be

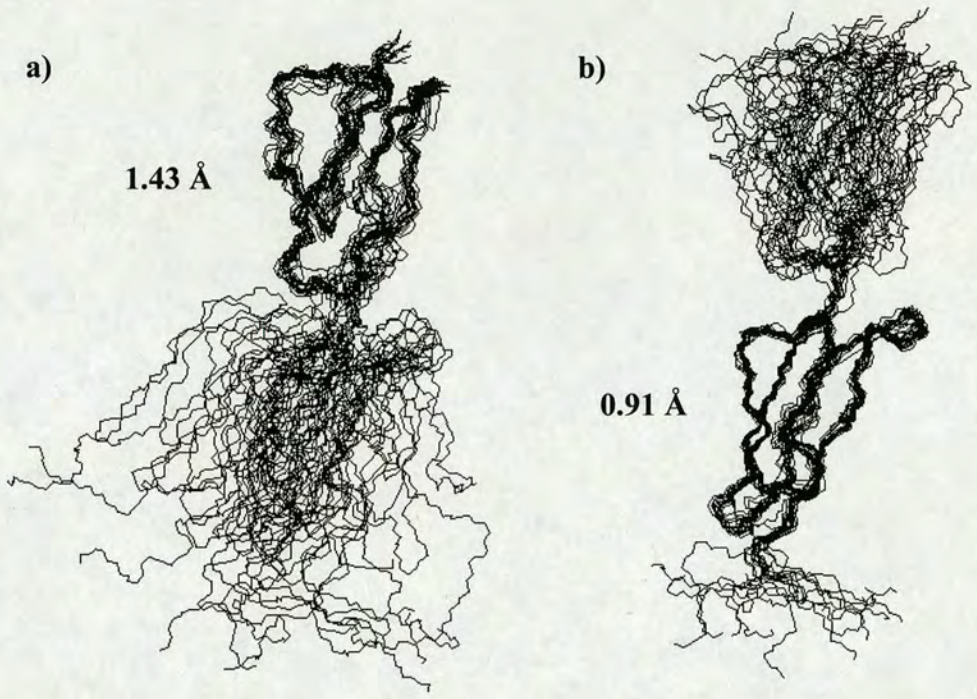


Figure 7.3: Lowest energy (18/100) structures obtained using unambiguous and semi-ambiguous HNH- and HCH_3-NOESY restraints. a) Structures overlaid on module 2 (residues 65-122) b) structures overlaid on module 1 (residues 2-60). The numbers indicate backbone rmsds.

employed to filter the final set of restraints to be added – those of the peaks with completely ambiguous assignments.

The inclusion of the completely ambiguous restraint information results in a much improved structure for both modules. Subsequent structure calculations were performed following this round to further refine the ambiguous restraint information. In total, 3 further rounds of calculations were performed to give an ensemble of structures shown in Figure 7.4. The structure calculation statistics of the ensemble are given in Tables 7.1 and 7.2 and the energy profile in Figure 7.5.

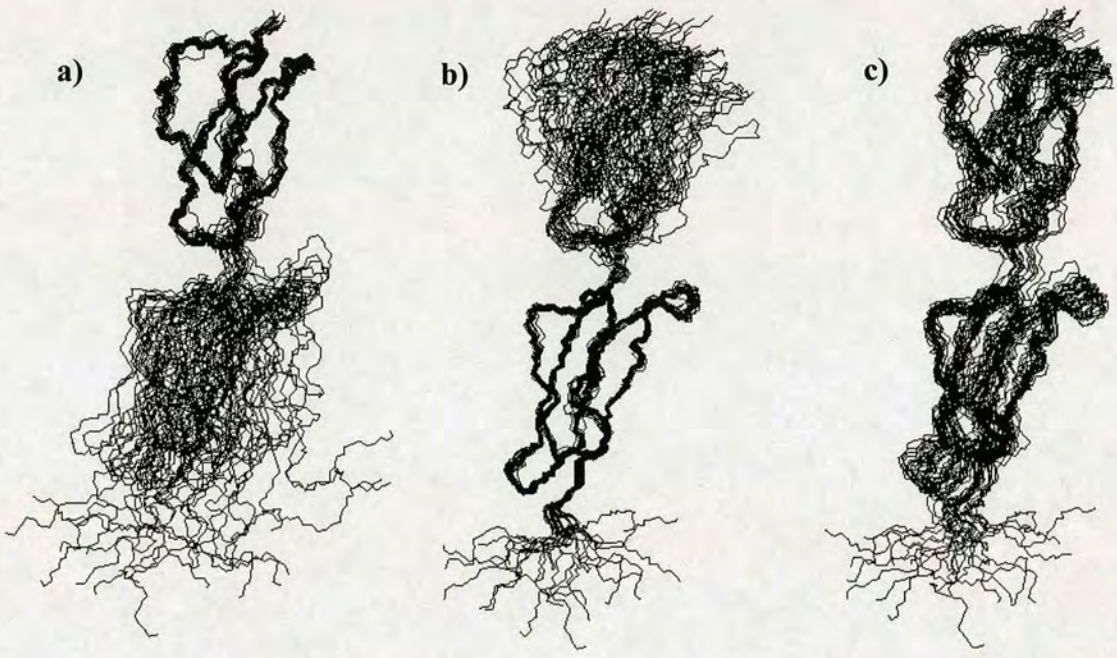


Figure 7.4: Lowest energy structure ensembles (23/100) calculated using all HNH- and HCH_3-NOESY restraint data overlaid on, a), module 1, b), module 2 and, c), on both modules.

NOE type	**No. restraints[a]**
Intra-residue	578
Sequential	677
Short-range (2-4 Å)	406
Long-range (>4 Å)	886
Ambiguous	771

	Rmsds (Å)		
Atoms	Module 1	Module 2	Both modules
Backbone	0.89	0.60	1.90
Heavy	1.39	1.13	2.24

Violations	
NOE (>0.3 Å)	3

Tables 7.1, 7.2 & 7.3: NOE and structure statistics for structure ensemble (23/100) calculated using all HNH- and HCH_3-NOESY restraint data. [a] These are the numbers of both unambiguous and semi-ambiguous unique restraints

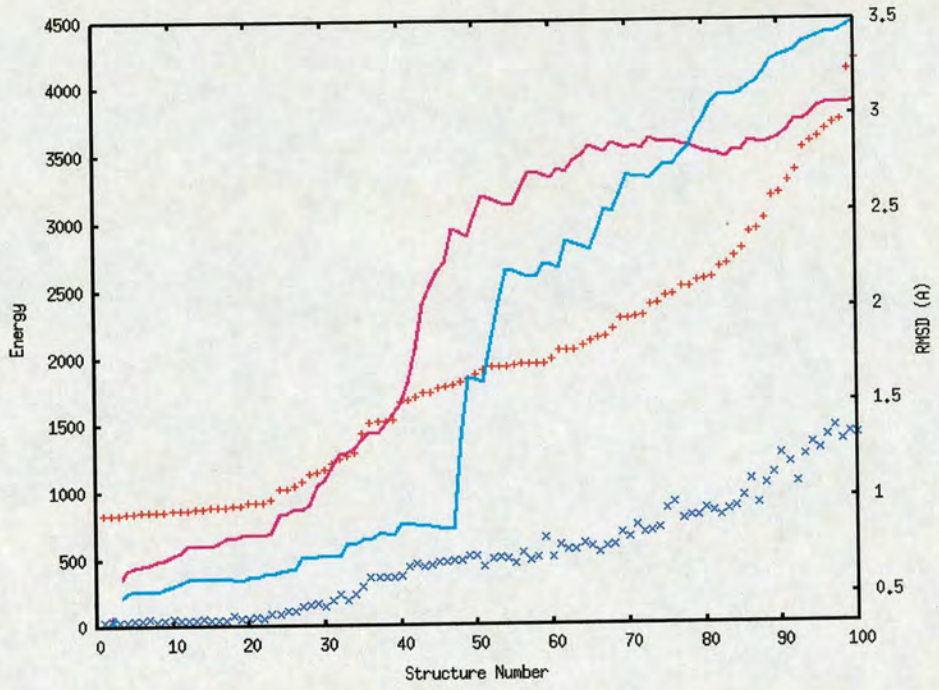


Figure 7.5: Plot of total potential energy (red) and NOE energy (blue) for the 100 structures calculated with all HNH- and HCH_3-NOESY restraints. Backbone rmsds of module 1 (magenta) and module 2 (cyan) are also plotted.

The converged structures show good agreement when overlaid on individual modules and, though somewhat poorer, the overlay of both modules is also acceptable. Further filtering and structure calculation steps did not improve upon this ensemble and, therefore, this was judged to be the conclusion of the structure determination with regard to the information available from the HNH-NOESY and HCH_3-NOESY spectra.

Comparison of these structures with those obtained by Jenkins shows a favourable agreement. In the calculations performed by Jenkins, a total of 4436 NOE distance restraints were used, 3541 of which were unambiguous and 895 unambiguous. The calculation procedures differed somewhat as the data used by Jenkins et al were principally derived from the HCH-NOESY spectrum in the first instance, whereas these data are not yet included in the calculations presented here. Also, Jenkins used the

assignment program ANSIG to assign the spectra resulting in only unambiguous or ambiguous restraint classes (no semi-ambiguous). The relevant structure statistics for this comparison are given in Table 7.4. It should be noted that these statistics are not for the final, published ensemble but refer to the ensemble obtained prior to the inclusion of residual dipolar coupling (RDC) restraints i.e. the structures obtained using only NOE-derived experimental restraint data.

	Rmsd's (Å)		
Atoms	Module 1	Module 2	Both modules
Backbone	1.091	0.764	2.716
Heavy	1.593	1.199	3.015

Table 7.4: Rmsds for the structure ensemble (26 out of 100) of Jenkins generated using NOE data (prior to RDC refinement)

The structures generated using the total HNH- and HCH_3-NOESY (unambiguous, semi- and completely ambiguous) restraints compare well with these statistics - the NOE statistics for these structures and the relevant structural data are given previously in Table 7.2. The total amount of information acquired is greater for the structures of Jenkins but a comparison of the structures themselves show that the Protocol, while giving slightly fewer converged structures overall, produces an ensemble with much better rmsds. As was noted with BlgB, and will be seen later (Tables 7.8-7.10) this difference can be attributed to the Protocol providing a greater number of restraints of high structural significance.

It is still possible to improve the quality of the Protocol-derived structures, however, by including the information which remains unused in the HCH-NOESY. As with BlgB, only the HNH- and HCH_3-NOESY restraints have been used to arrive at this point and no quantitative information from the HCH-NOESY has yet been obtained. In particular, the orientation of the sidechains would be expected to benefit from this further

refinement. Following the procedure used for BlgB, four regions of the HCH-NOESY spectrum were selected for automated peak picking, resulting in 2234 peaks which were converted into distance restraints using the 'Shift Match Distance Constraints' function of analysis. These completely ambiguous restraints were then exported and filtered using the structure ensemble obtained previously prior to performing a further calculation. In total, 1359 unique restraints were used.

Progressive filtering of these restraints in three rounds of calculations gave the final NOE-derived structure ensemble, the statistics for which are given in Table 7.5.

	Rmsd's (Å)		
Atoms	Module 1	Module 2	Both modules
Backbone	0.85	0.54	1.49
Heavy	1.21	1.02	1.90

Table 7.5: Structure statistics for the structure ensemble (33/100) calculated using HNH-, HCH_3- and HCH-NOESY restraints

The additional carbon restraints result in an increased number of converged structures (33) and improved rmsd's as shown in Figure 7.6. It is worth noting that the difference in the rmsds of backbone atoms in the structures with and without the additional carbon restraints are only marginal. A much more substantial reduction is seen in the rmsd's of the sidechain atoms. This is as might be expected since the HNH- and HCH_3-NOESY data will largely define only backbone positions and the positions of chain termini (for those sidechains which have amine or methyl groups). The carbon data will contain restraints which define the position of the entire sidechain thus leading to sidechain positions which are much more precisely defined.

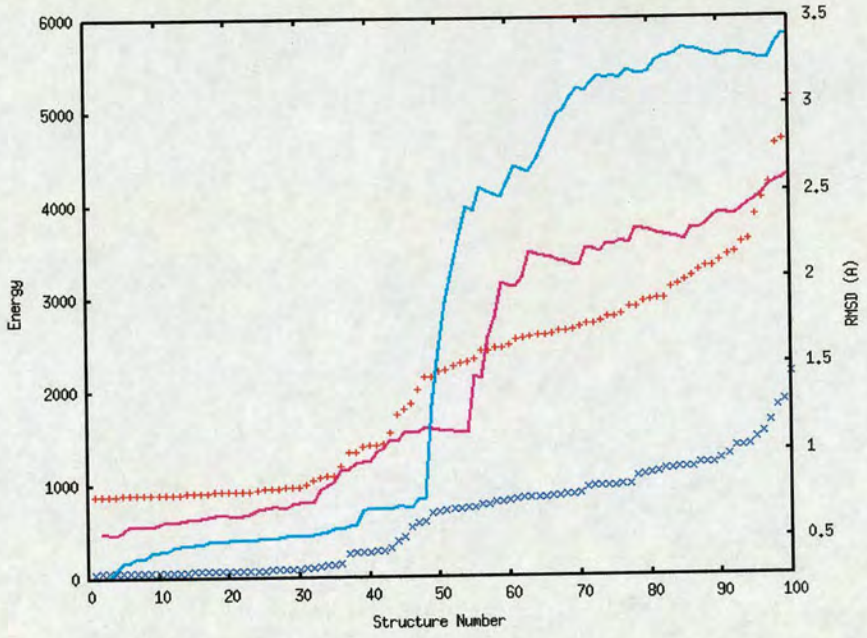


Figure 7.6: Plot of total potential energy (red) and NOE energy (blue) versus structure number for ensemble of structures calculated using HNH-, HCH_3- and HCH-NOESY spectra restraints. Backbone rmsds for module 1 (magenta) and module 2 (cyan) are also plotted.

Overlays of the closest to mean structures obtained from this ensemble and the final (RDC refined) ensemble of Jenkins et al, given in Figure 7.7 show good agreement with backbone rmsds of 1.2 Å and 1.1 Å for modules 1 and 2 respectively. Overlaying both modules gives a larger difference of 1.95 Å. This is caused by the differing intermodular orientations of the two structures. This difference can be attributed to the use of RDCs in the refinement of the structures calculated by Jenkins et al. As RDCs are inherently long-range restraints, they are particularly useful in improving the definition of the intermodular orientation. Inclusion of the RDC restraints used by Jenkins – 62 $^1D_{NH}$, 65 $^1D_{NC'}$, 41 $^1D_{C\alpha C'}$, 24 $1D_{C\alpha H\alpha}$ - in a final structure refinement step of the calculations presented here gives a much better agreement as shown in Figure 7.7 and Table 7.6 with a backbone rmsd of 1.1 Å between the closest to mean structures (overlaid on both modules) and also increases the number of structures which converge on the lowest energy conformation.

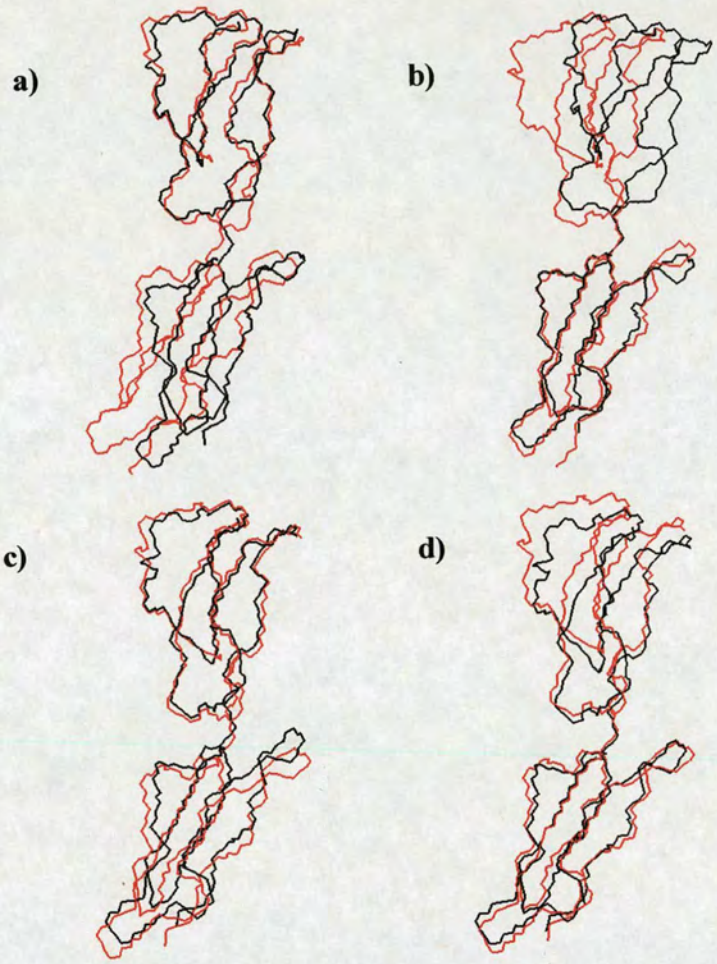


Figure 7.7: Comparison of closest to mean (RDC refined) structure obtained by Jenkins et al (red) and that calculated here (black). a) & b) show overlays on module 1 and 2 respectively for the structure obtained without RDCs and c) & d) show the same for the structure obtained with RDCs.

	Angle (°)[a]		
Structure	**Tilt**	**Twist**	**Skew**
NOE only	24.50	-1.05	158.44
NOE + RDC	34.76 (38.42)	4.31 (3.66)	252.32 (249.85)

Table 7.6: Intermodular orientations for the NOE only and the NOE + RDC structures. [a] Numbers in parenthesis indicate the values for the structure of Jenkins et al.

Clearly, the RDC refinement gives structures with intermodular orientations which agree much more closely with those of the published structure than is found for those structures which were obtained without RDCs. A comparison of the structural statistics for this final structure ensemble with the ensemble calculated by Jenkins et al is given in Table 7.7.

	Rmsds (Å)[a]		
Atoms	Module 1	Module 2	Both modules
Backbone	0.85 (0.90)	0.54 (0.60)	0.982(1.04)
Heavy	1.19 (1.49)	1.00 (1.11)	1.40 (1.51)

Table 7.7: Rmsd's for final structure ensemble (50 out of 100) calculated usin HNH-, HCH_3- and HCH-NOESY restraints and RDC restraint data. [a] Numbers in parentheses are the values for the structure ensemble (40 out of 100) calculated by Jenkins et al.

7.5 Analysis of the Performance of the Protocol

The results of the previous section indicate that the Protocol performs somewhat more effectively than Method 1 with a good quality 'NOE-only' structure of C4BP~1,2 being obtained with fewer total restraints. However, as with BlgB, a more quantitative approach is necessary to obtain a precise measure of the improvement gained via the protocol. To this end, assignment of the HNH-NOESY and HCH_3-NOESY spectra was performed again using Method 1 but, in this case, a side-by-side comparison of the two methods was not performed. Instead the results of the structure calculation were used to perform a retrospective analysis.

The first and simplest measure is a comparison of the assignments obtained via Methods 1 and 2, indicating the degree to which the total assignment ambiguity has been reduced. Table 7.8 gives the results of comparing these two sets of data.

Spectrum	Method	Unambiguous	Semi-ambiguous	Ambiguous
HNH-NOESY	1	1188	946	363
	2	1308	803	386
HCH_3-NOESY	1	379	654	299
	2	498	531	303

Table 7.8: Comparison of number of unambiguous, ambiguous and semi-ambiguous NOEs obtained via Methods 1 and 2. These are the totals prior to removal of erroneous and duplicate restraints.

Method 2 shows a greater number of unambiguous assignments, as would be expected. As was observed in the case of BlgB, a greater number of ambiguous restraints is obtained corresponding to peaks whose assignment possibilities did not match with any of the peaks of the complementary spectrum (instances of case B).

A peak by peak comparison of the two sets of assignments was performed to give a more accurate assessment of the assignment difference; determining the number of cases of ambiguity elimination and ambiguity reduction. Prior to performing this analysis, erroneous and duplicate restraints were removed from the datasets. These figures are summarized in Table 7.9.

Spectrum	Ambiguity Elimination	Ambiguity Reduction
HNH-NOESY	170	140
HCH_3-NOESY	184	173

Table 7.9: Number of cases of ambiguity elimination and reduction in the HNH- and HCH_3-NOESY spectra

In total, for the HNH-NOESY spectrum, 317 peaks were 'less ambiguous' when assigned via the protocol than when assigned via the standard method and, for the HCH_3-edited NOESY spectrum, 366 peaks were 'less ambiguous'. When compared to the total number of peaks from each spectrum, these figures are relatively small but this is somewhat misleading since the ambiguity reduction events occur only within a subset of the total peaks i.e. those peaks arising from NOE's to aliphatic protons. If the ambiguity elimination and reduction events are expressed as a fraction of this subset of the total data, however, they can be seen to represent a substantial portion of these data – ca. 20% of the HNH-NOESY and ca. 30 % of the HCH_3-NOESY data.

The composition of these events is given in Table 7.10. As shown here, of the assignments which led to satisfied restraints, the greater portion of these are the more structurally significant short- or long-range NOEs.

Spectrum	Event	Sequential	Short	Long
HNH-NOESY	elimination	25	56	89
	reduction	26	47	67
HCH_3-NNOESY	elimination	7	67	110
	reduction	26	80	67

Table 7.10: Breakdown of NOE types of ambiguity reduction events

A more precise measure of the importance of these restraints is obtained via QUEEN analysis of the Method 2 restraints. Figure 7.8 gives the relevant output information arising from the QUEEN analysis of the correct unambiguous restraints used to calculate the structures in Figure 7.4. Those restraints which have been identified as ambiguity elimination events are highlighted.

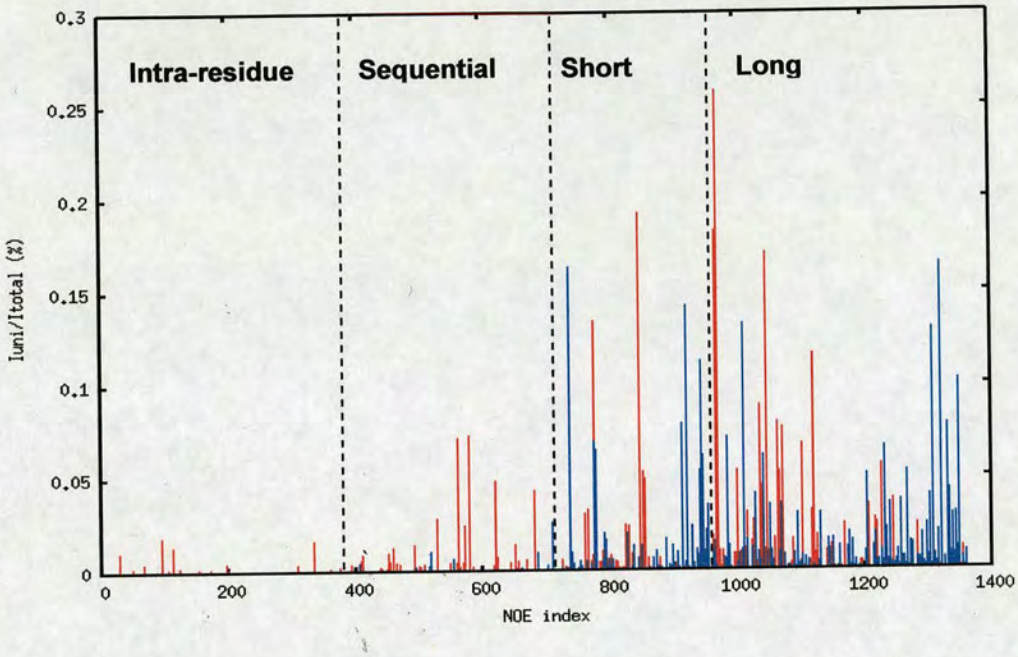


Figure 7.8: Plot of I_{uni}/I_{total} values for unambiguous HNH-NOESYand HCH_3-NOESY restraints. The blue lines indicate those restraints which are cases of ambiguity elimination.

As these analyses show, similar observations can be made as were seen for BlgB i.e. not only does the Protocol reduce the total assignment ambiguity for a spectrum, the restraints which these produce exist largely in the short- and long-range categories of NOE information and are found to be among the most unique pieces of information present in the dataset.

Once identified, the effect of the less ambiguous restraints upon the quality of structures obtained using these restraints can be assessed by replacing the restraints of interest in the restraint tables produced via Method 2 with their counterparts arising from assignment via Method 1. Obviously this will result in an overall increase in the total ambiguity present in the dataset. Figure 7.9 shows the results of calculations performed with these two differing sets of restraints.

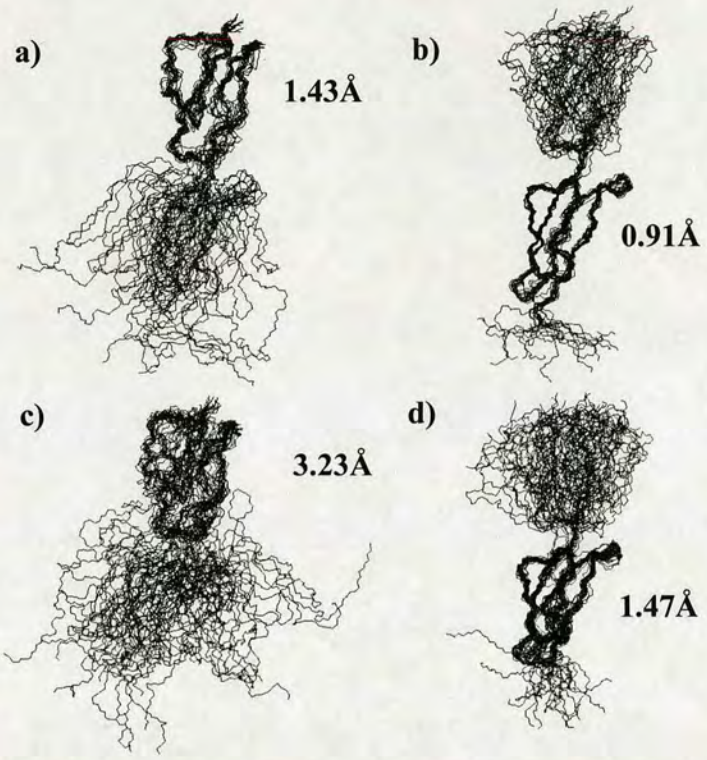


Figure 7.9: Lowest energy structures (20/100) overlaid on module 1, a), and module 2, b), obtained using Method 2 restraints. c) and d) show the same overlays but for structures calculated with the ambiguity reduction and elimination events replaced by Method 1 assignments. Numbers indicate the backbone rmsds for the overlaid residues

The increase in ambiguity clearly has a noticeable effect upon the overall structure quality resulting in much poorer rmsds for both modules. Importantly, for module 1, the structures cease to converge on a single conformation giving rmsds which are correspondingly much worse. It should be noted that this 'composite' restraint set is obviously artificial and, therefore, the results given above do not necessarily represent the output which would be obtained at any point in calculating these structures using spectra assigned via Method 1. However, it still provides a rough assessment of the impact of the greater information content of restraints obtained via Method 2, in the absence of a side-by-side comparison. Thus, the improvements seen for assignment of BlgB – greater number of unambiguous restraints and overall reduced ambiguity of

ambiguous restraints – which delivered improved structures more quickly are also seen for C4BP~1,2.

Furthermore, as was also seen in the case of BlgB, assigning via the protocol provides an additional safeguard against incorrect assignments. Comparing the initial restraint tables provided by each method with the final structure shows that the reduction in the number of incorrect assignments is quite significant. Of the 1567 unambiguous restraints obtained via Method 1, 161 (ca. 10%) were found to be incorrect whereas, for Method 2, 93 of the 1806 (ca. 5%) unambiguous assignments were incorrect. Thus Method 2 seems to perform consistently with similar increases in the amount of correct, structurally significant, unambiguous information and decreases in the number of incorrect assignments as were seen for BlgB.

In total, the time required to obtain a structure for C4BP~1,2 from the beginning of NOESY assignment was of the order of ca. 4 weeks. In contrast, the time spent by Jenkins in obtaining the same structure was ca. 24 weeks.

Chapter 8

Conclusions

Four complementary 3D NOESY experiments (CNH-, NNH-, CCH- and CCH_3-NOESY) employing heteronuclear labeling in both indirect dimensions have been designed and implemented for use as assignment aids in assigning 'standard' NOESY spectra (HNH-, HCH- and HCH_3-NOESY). Of the four, two spectra (CNH- and CCH_3-NOESY) were deemed to be most suitable for use as assignment aids on the basis of signal density and spectral resolution. A protocol for efficient and effective incorporation of these spectra into the NOESY spectra assignment procedure was designed, implemented and tested in the Analysis software package.

Assignment of the 3D HNH- and HCH_3-NOESY spectra of BlgB (18 kDa) was performed with and without the use of the complementary experiments to test the performance of the Protocol. It was found that assignments made via the Protocol were less ambiguous overall and yielded a restraint set with significantly improved information content. This, coupled with increased accuracy of assignments, delivered structures of better quality more quickly than was the case when the complementary spectra were not used. Inclusion of a subset of restraints present in the HCH-NOESY spectrum was achieved in a simple and efficient manner via automated peak selection and the use of the structure ensemble provided by the HNH- and HCH_3-NOESY spectra to filter the ambiguous assignments. This procedure eliminated the need for assignment of the 3D HCH-NOESY data – the most time-consuming step in NMR-based structure determination of proteins.

As a preliminary step for this work, near-complete resonance assignment was achieved for the protein β-lactoglobulin B and the low pH solution structure of this protein was solved. The structure ensemble obtained showed excellent convergence with low backbone rmsds and low total energies. The structural validity was assessed via quality checks performed with PROCHECK and WHATIF which showed no unusual results. The low pH solution structure of BlgB closely resembles that of BlgA; potential differences of interest due to changes in two amino acids were undiscernable due to flexibility in the relevant area.

In order to further test the Protocol, the HNH- and HCH_3-NOESY spectra of the double CCP module construct C4BP~1,2 (15 kDa) were assigned. Comparing the resulting restraint sets with those obtained without the use of the complementary spectra showed an overall decrease in ambiguity for the Protocol dataset along with similar increases in the structural significance and information content of restraints as were observed in the case of BlgB. The solution structure of C4BP~1,2 was solved using Protocol-derived HNH- and HCH_3-NOESY restraints and completely ambiguous set of HCH-NOESY restraints. Following refinement using RDCs, the resulting structures agreed well with the structure solved previously (Jenkins et al, 2005) and showed somewhat improved convergence with slightly lower backbone rmsds and an increased number of converged structures.

The use of the Protocol outlined here requires that three additional 3D experiments be acquired compared to a 'standard' assignment method thus presenting a potential disadvantage in the amount of time which must be spent acquiring spectra. The significance of this problem is far from prohibitive, however, as these time requirements can easily be absorbed into other time-intensive procedures such as resonance assignment. Furthermore, as has been shown, the potential time gains are much more substantial, reducing the time required for structure determination of medium size (~20 kDa) double labeled proteins from months to weeks.

Overall, the Protocol developed here has been shown to be effective at providing good quality structures in relatively short time. The Protocol demonstrated robustness and flexibility, performing consistently with both the protein samples used in the investigation; providing similar decreases in total assignment ambiguity coupled with an increase in the numbers of structurally significant restraints. For both BlgB and C4BP~1,2, the structures obtained using the HNH- and HCH_3-NOESY data were of good quality with low (<1.0 Å) backbone rmsds. Further improvements, in particular to the position of sidechains, were obtained on including HCH-NOESY data. By employing this Protocol, the total time spent assigning spectra and calculating and refining structures can be greatly reduced without compromising the quality of the NMR structures of proteins.

Bibliography

BioMagResBank (BMRB) web site: http://www.bmrb.wisc.edu

W. P. Aue, E. Bartholdi & R. R. Ernst. Two-Dimensional Spectroscopy. Application to Nuclear Magnetic Resonance. *J. Chem. Phys.*, 64:2229-2246, 1976.

A. Bax. Two-Dimensional NMR and Protein Structure. *Annu. Rev. Biochem.*, 58:223-256, 1989.

A. Bax & S. Pochapsky, Optimized Recording of Heteronuclear Multidimensional NMR Spectra Using Pulsed Field Gradients. *J. Magn. Reson.*, 99:638-643, 1992.

A. M. Blom, L. Kask & B. Dahlbaeck. Structural Requirements for the Complement Regulatory Activities of C4BP. *J. Biol. Chem.*, 276:27136-27144, 2001.

R. Boelens, M. Burgering, R. H. Fogh & R. Kaptein. Time-Saving Methods for Heteronuclear Multidimensional NMR of (^{13}C, ^{15}N) Doubly Labelled Proteins, *J. Biomol. NMR*, 4:201-213, 1994.

W. Boucher & University of Cambridge Department of Biochemistry. Azara, v2.7, 1993.

S. Brownlow, J. H. M. Cabral, R. Cooper, D. R. Flower, S. J. Yewdall, I. Polikarpov, A. C. T. North & L. Sawyer. Bovine β-Lactoglobulin at 1.8 Å Resolution – Still an Enigmatic Lipocalin. *Structure*, 5:481-495, 1997.

A. T. Brunger. X-PLOR, version 3.1: A System for X-ray Crystallography and NMR. *Yale University Press*, New Haven, 1992.

A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice & T. Simonson. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta. Cryst.* D54:905-921, 1998.

B. Brutscher. Combined Frequency- and Time-Domain NMR Spectroscopy. Application to Fast Protein Resonance Assignment. *J. Biomol. NMR*. 29:57-64, 2004.

K. A. Carpenter & F. Ni. A Heteronuclear 3D NMR Experiment Which Allows Detection of All NOE's for a Molecule While Providing ^{15}N Frequency Separation of Degenerate Amide Protons. *J. Magn. Reson.* 99:192-197, 1992.

J. Cavanagh, A. G. PalmerIII, P. E. Wright & M. Rance. Sensitivity Improvement in Proton-Detected Two-Dimensional Heteronuclear Relay Spectroscopy. *J. Magn. Reson.*, 91:429-436, 1991.

C. Dalvitt, G. Bovermann, K. Memmert & M. Zurini, Sequential Assignment of β-sheet Regions of Proteins via a 2D ^{15}N-^{1}H HSQC with a Homonuclear-Relay-NOESY Step. *J. Magn. Reson.* 96:174-180, 1992.

B. Dahlbaeck, C. A. Smith & H. J. Mueller-Eberhard. Visulaization of Human C4b-Binding Protein and its Complexes with Vitamin K-Dependant Protein S and Complement Protein C4b. *Proc. Natl. Acad. Sci. U.S.A.*, 80:3461-3465, 1983.

R. Diamond. Coordinate-Based Cluster-Analysis. *Acta. Cryst.* D51:127-135, 1995.

T. Diercks, M. Coles & H. Kessler. An Efficient Strategy for Assignment of Cross-Peaks in 3D Heteronuclear NOESY Experiments. *J. Biomol. NMR*, 15:177-180, 1999.

L. Emsley, G. Bodenhausen. Gaussian Pulse Cascades: New Analytical Functions for Rectangular Selective Inversion and In-Phase Excitation in NMR. *Chem. Phys. Lett.*, 165:469-476, 1990.

B. T. Farmer II. Simultaneous [^{13}C, ^{15}N]-HMQC, A Pseudo-Triple-Resonance Experiment, *J. Magn. Reson.*, 93:635-641, 1991.

B. T. Farmer & L. Mueller. Simultaneous Acquisition of [^{13}C,^{15}N]- and [^{15}N,^{15}N]-Separated 4D Gradient-Enhanced NOESY Spectra in Proteins. *J. Biomol. NMR*, 4:673-687, 1994.

S. W. Fesik & E. R. P. Zuiderweg. Heteronuclear 3D NMR Spectroscopy. A Strategy for the Simplification of 2D NMR Spectra. *J. Magn. Reson.* 78:588-593, 1988.

R. Freeman & E. Kupce. New Methods for Fast Multidimensional NMR. *J. Biomol. NMR*. 27:101-113, 2003

T. Frienkel, C. Bauer, M. D. Carr, B. Birdsall & J. Feeney, HMQC-NOESY-HMQC, a 3D NMR Experiment Which Allows Detection of NOE's Between Protons with Overlapping Signals. *J. Magn. Reson.*, 90:420-425, 1990.

L. Frydman, T. Scherf & A. Lupulescu. The Acquisition of Multidimensional NMR Spectra Within a Single Scan. *Proc. Natl. Acad. Sci. U.S.A.,* 99:15858-15862, 2002.

R. Fu & G. Bodenhausen. Ultra-Broadband Decoupling. *J. Magn. Reson.* A, 117:324-325, 1995.
A. Grishaev & M. Llinas. CLOUDS, a Protocol for Deriving a Molecular Proton Density via NMR. *Proc. Natl. Acad. Sci. U.S.A.*, 99:6707-6712, 2002a.

A. Grishaev & M. Llinas. Protein Structure Elucidation from NMR Proton Densities. *Proc. Natl. Acad. Sci. U.S.A.,*. 99:6713-6718, 2002b.

S. Grzesiek, J. Anglister & A. Bax. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in ^{13}C/^{15}N-Enriched Proteins by Isotropic Mixing of ^{13}C-Magnetization. *J. Magn. Reson. B*, 101:114-119, 1993.

S. Grzesiek & A. Bax. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J. Am. Chem. Soc.*, 114:6291-6293, 1992.

S. Grzesiek & A. Bax. The Importance of Not Saturating H_2O in Protein NMR. Application to Sensitivity Enhancement and NOE Measurements. *J. Am. Chem. Soc.*, 115:12593-12594, 1993a.

S. Grzesiek & A. Bax. Amino Acid Type Detemination in the Sequential Assignment Procedure of Uniformly ^{13}C/^{15}N-Enriched Proteins. *J. Biomol. NMR*, 3:185-204, 1993b.

S. Grzesiek, P. Wingfield, S. Stahl, J. D. Kaufamn & A. Bax. Four-Dimensional ^{15}N-Separated NOESY of Slowly Tumbling Perdeuterated ^{15}N-Enriched Proteins. Application to HIV-1 Nef. *J. Am. Chem. Soc.*, 117:9594-9595, 1995.

P. Guentert. Automated NMR Protein Structure Calculation. *Prog. NMR Spec.*, 43:105-125, 2003.

S.F. Gull, G. J. Daniell. Image Reconstruction from Incomplete and Noisy Data. *Nature*, 272:686-272, 1978.

S. G. Hambling, A. S. McAlpine & L. Sawyer. Advanced Dairy Chemistry: 1. Proteins. *Elsevier Applied Science*, Amsterdam, 1992.

T. Herrmann, P. Guentert & K. Wuethrich. Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *J. Mol. Biol.* 319:209-227, 2002.

J. P. Hill et al. Macromolecular Interactions in Food Technology. *American Chemical Society*, Washington, DC, 1996.

P. J. Hore. Nuclear Magnetic Resonance, OCP 32, *Oxford University Press*, Oxford, 1995.

P. J. Hore, J. A. Jones & S. Wimperis. NMR: The Toolkit, OCP 92, *Oxford University Press*, Oxford, 2000.

D. I. Hoult & R. E. Richards. Critical Factors in the Design of Sensitive High Resolution Nuclear Magnetic Resonance Spectrometers. *Proc. R. Soc. Lond. A*, 344:311-340, 1975.

M. Ikura, L. E. Kay, R. Tschudin & A. Bax, 3D NOESY-HMQC Spectroscopy of a ^{13}C-Labelled Protein. *J. Magn. Reson.*, 86:204-209, 1990.

J. Jeener. Ampere Summer School, Yugoslavia, 1971.

H. T. Jenkins, L. Mark, G. Ball, J. Persson, G. Lindahl, D. Uhrin, A. M. Blom & P. N. Barlow. Human C4b-binding Protein, Structural Basis for Interaction with Streptococcal M Protein, a Major Bacterial Virulence Factor. *J. Biol. Chem.*, 281:3690-3697, 2006.

R. Jerala & G. S. Rule. A 3D ^{1}H, ^{15}N and ^{13}C NOESY Correlating Experiment. *J. Magn. Reson. B*, 108:294-298, 1995.

Y. S. Jung & M. Zweckstetter. Mars – Robust Automatic Backbone Assignment of Proteins. *J. Biomol. NMR*, 30:11-23, 2004.

L. E. Kay, G. M. Clore, A. Bax, A. M. Gronenborn. Four-Dimensional Heteronuclear Triple-Resonance NMR Spectroscopy of Interleukin-1b in Solution. *Science*, 249:411-414, 1990.

L. E. Kay, P. Keifer & T. Saarinen. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J. Am. Chem. Soc.*, 114:10663-10665, 1992.

L. E. Kay, M. Wittekind & A. McCoy. 4D NMR Triple-Resonance Experiments for Assignment of Protein Backbone Nuclei Using Shared Constant-Time Evolution Periods. *J. Magn. Reson.*, 98:443-456, 1992.

L. E. Kay, G. Y. Xu, A. U. Singer, D. R. Muhandiram & J. D. Forman-Kay. A Gradient Enhanced HCCH-TOCSY Experiment for Recording Side Chain ^{1}H and ^{13}C Correlations in H_2O samples of Proteins. *J. Magn. Reson. B*, 101:333-337, 1993.

H. Kessler, M. Gehrke & C. Griesinger, 2D NMR Spectroscopy: Background and Overview of Experiments, *Angew. Chem. Int. Ed. Engl.*, 27:190-536, 1988.

S. Kim & T. Szyperski. GFT NMR a New Approach to Rapidly Obtain Precise High-Dimensional NMR Spectral Information. *J. Am. Chem. Soc.* 125:1385-1393, 2003.

R. Koradi, M. Billeter & K. Wuethrich. MOLMOL: A Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graph.*, 14: 51-55, 1996.

P. J. Kraulis. A Program for the Assignment of Protein ^{1}H 2D NMR Spectra by Interactive Graphics. *J. Magn. Reson.*, 24:627-633, 1989.

E. Kupce & R. Freeman. Adiabatic Pulses for Wideban inversion and Broadband Decoupling. *J. Magn. Reson.* A, 115:273-276, 1995.

E. Kupce & R. Freeman. Two-Dimensional Hadamard Spectroscopy. *J. Magn. Reson.* 162:300-310, 2003

E. Kupce & R. Freeman. Projection-Reconstruction for Speeding Up Multidimensional NMR Spectroscopy. *J. Am. Chem. Soc.* 126:6429-6440, 2004.

J. Kuszewski, C. D. Schwieters, D. S. Garrett, R. A. Byrd, N. Tjandra & G. M. Clore. Completely Automated, Highly Error-Tolerant Macromolecular Structure Determination from Multidimensional Nuclear Overhauser Enhancement Spectra and Chemical Shift Assignments. *J. Am. Chem. Soc.*, 126:6258-6273, 2004.

R. A. Laskowski, J. A. C. Rullman, M. W. MacArthur, R. Kaptein & J. M. Thornton. AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR*, 8:477-486, 1996.

R. A. Laskowski, M. W. MacArthur, D. S. Moss & J. M. Thornton. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.*, 26:283-291, 1993.

J. P. Linge & M. Nilges. Influence of Non-Bonded Parameters on the Qulaity of NMR Structures: A New Forcefield for NMR Structure Calculation. *J. Biomol. NMR*, 13:51-59, 1999.

T. M. Logan, E. T. Olejniczak, R. X. Xu & S. W. Fesik. A General Method for Assigning NMR Spectra of Denatured Proteins Using 3D HC(CO)NH-TOCSY Experiments. *J. Biomol. NMR*, 3:225-231, 1993.

F Loehr & H. Rueterjans. A New Triple-Resonance Experiment for the Sequential Assignment of Backbone Resonances in Proteins. *J. Biomol. NMR*, 6:189-197, 1995.

M. Madrid, E. Llinas & M. Llinas. Model Independent Refinement of Interproton Distances Generated from ^{1}H NMR Overhauser Intensities. *J. Magn. Reson.*, 93:329-346, 1991.

A. Majumdar & E. R. P. Zuiderweg. Improved ^{13}C-Resolved HSQC-NOESY Spectra in H_2O Using Pulsed Field Gradients. *J. Magn. Reson.*, 102:242-244, 1993.

D. Marion, L. E. Kay, S. W. Sparks, D. A. Torchia & A. Bax. Three-Dimensional Heteronuclear NMR of ^{15}N-labelled Proteins. *J. Am. Chem. Soc.*, 111:1515-1517, 1989.

J. E. Masse, R. Keller & K. Pervushin. SideLink: Automated Side-Chain Assignment of Biopolymers from NMR Data by Relative-Hypothesis-Prioritization-Based Simulated Logic. *J. Magn. Reson.*, 181:45-67, 2006.

R. C. Morshauser & E. R. P. Zuiderweg. High-Resolution Four-Dimensional HMQC-NOESY-HSQC Spectroscopy. *J. Magn. Reso*n., 139:232-239, 1999.

G. A. Morris & R. Freeman. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *J. Am. Chem. Soc.*, 101:760-762, 1979.

H. N. B. Moseley & G. T. Montelione. Automated Analysis of NMR Assignments and Structures. *Curr. Opin. Struct. Biol.* 9:635-642, 1999.

D. R. Muhandiram, N. A. Farrow, G. Xu, S. H. Smallcombe & L. E. Kay. A Gradient ^{13}C NOESY-HSQC Experiment for Recording NOESY spectra of ^{13}C-Labelled Proteins Dissolved in H_2O. *J. Magn. Reson.*, 102:317-321, 1993.

D. R. Muhandiram, G. Y. Xu & L. E. Kay. *J. Biomol. NMR*, 3:463-470, 1993.

C. Mumenthaler, P. Guentert, W. Braun, K. Wuethrich. Auotomated Combined Assignment of NOESY spectra and Three-dimensional Protein Structure Determination. *J. Biomol. NMR*. 10:351-362, 1997.

S. B. Nabuurs, A. J. Nederveen, W. Vranken, J. F. Doreleijers, A. M. J. J. Bonvin, G. W. Vuister, G. Vriend, C. A. E. M. Spronk. DRESS: a Database of Refined Solution NMR Structures. *Proteins* 55:483-486, 2004.

S. B. Nabuurs, C. A. E. M. Spronk, E. Krieger, H. Maassen, G. Vriend & G. W. Vuister. Quantitative Evaluation of Experimental NMR Restraints. *J. Am. Chem. Soc.* 125:12026-12034, 2003.

S. B. Nabuurs, E. Krieger, C. A. E. M. Spronk, A. J. Nederveen, G. Vriend & G. W. Vuister. Definition of a new Per-Residue Based Quality Parameter. *J. Biomol. NMR*, 33:123-134, 2005.

A. J. Nederveen, J. F. Doreleijers, W. Vranken, Z. Miller, C. A. E. M. Spronk, S. B. Nabuurs, P. Guentert, M. Livny, J. L. Markley, M. Nilges, E. L. Ulrich, R. Kaptein & A. M. J. J. Bonvin.

RECOORD: a REcalculated COORdinates Database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* 59:662-672, 2005.

D. Neuhaus & M. P. Williamson, The Nuclear Overhauser Effect in Structural and Conformational Analysis, 2nd edn, *Wiley-VCH*, 2000.

M. Nilges. Structure Calculation from NMR Data. *Curr. Opin. Struct. Biol.*, 6:617-623, 1996.

M. Nilges, M. J. Macias, S. O'Donoghue & H. Oschkinat. Automated NOESY Interpretation with Ambiguous Distance Restraints: The Refined NMR Solution Structure of the Pleckstrin Homology Domain from β-Spectrin. *J. Mol. Biol.* 269:408-422, 1997.

M. Nilges & S. I. O'Donoghue. Ambiguous NOEs and Automated NOE Assignment. *Prog. NMR Spec.*, 32:107-139, 1998.

K. M. G. Oliviera, V. L. Valente-Mesquita, M. M. Botelho, L. Sawyer, S. T. Ferreira & I. Polikarpov. Crystal Structures of Bovine β-lactoglobulin in the Orthorhombic Space Group $C222_1$. *Eur. J. Biochem.* 268:477-483, 2001.

S. M. Pascal, D. R. Muhandiram, T. Yamazaki, J. D. Forman-Kay & L. E. Kay, Simultaneous Acquisition of ^{15}N- and ^{13}C-Edited NOE Spectra of Proteins Dissolved in H_2O, *J. Magn. Reson.*, 103:197-201, 1994.

M. Piotto, V. Saudek & V. Sklenar, Gradient Tailored Excitation for Single Quantum NMR-Spectroscopy of Aqueous Solutions. *J. Biomol. NMR*, 2:661-665, 1992.

B. Y. Qin, M. C. Bewley, L. K. Creamer, H. M. Baker, E. N. Baker, G. B. Jameson. Structural Basis of the Tanford Transition of Bovine β-Lactoglobulin. *Biochemistry*, 37:14014-14023, 1998.

B. Y. Qin, M. C. Bewley, L. K. Creamer, H. M. Baker, E. N. Baker, G. B. Jameson. Functional Implications of Structural Differences between Variants A and B of Bovine β-Lactoglobulin. *Protein Sci.*, 8: 75-83, 1999.

G. N. Ramachandran & V. Sasiskharan. Conformations of Polypeptides and Proteins. *Adv. Protein Chem.* 23:283-437, 1968.

J. K. M. Saunders & B. Hunter, Modern NMR Spectroscopy, 2nd edn, Oxford University Press, 1993.

A. J. Shaka, P. B. Barker & R. Freeman. Computer Optimized Decoupling Scheme for Wide-Band Applications and Low-Level Operation. *J. Magn. Reson.*, 64:547-551, 1985.

S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue & J. Staunton. Maximum Entropy Signal Processing in Practical NMR Spectroscopy. *Nature*, 311:446-447, 1984.

J. P. Simorre, B. Brutscher, M. S. Caffrey & D. Marion. Assignment of NMR Spectra of Proteins Using Triple-Resonance Two-Dimensional Experiments. *J. Biomol. NMR*, 4:325-334, 1994.

I. Solomon. Relaxation Processes in a System of Two Spins. *Phys. Rev.*, 99:559-565, 1955.

O. W. Sørensen, Aspects and Prospects of Multidimensional Time-Domain Spectroscopy, *J. Magn. Reson.*, 89:210-216, 1990.

O. W. Sørensen, G. W. Eich, M. H. Levitt, G. Bodenhausen & R. R. Ernst. Product Operator Formalism for the Description of NMR Pulse Experiments. *Progr. NMR Spec.*, 16:163-192, 1983.

C. A. E. M. Spronk, S. B. Nabuurs, E. Krieger, G. Vriend & G. W. Vuister. Validation of Protein Structures Derived by NMR Spectroscopy. *Prog. NMR. Spec.*, 45:315-337, 2004.

T. Szyperski, G. Wider, J. H. Bushweller & K. Wuethrich. Reduced Dimensionality in Triple-Resonance NMR Experiments. *J. Am. Chem. Soc.* 115:9307-9308, 1993a.

T. Szyperski, G. Wider, J. H. Bushweller & K. Wuethrich. 3D ^{13}C-^{15}N-Heteronuclear Two-Spin Coherence Spectroscopy for Polypeptide Backbone Assignments in ^{13}C-^{15}N-Double-Labeled Proteins. *J. Biomol. NMR*, 3:127-132, 1993b.

C. Tanford, L. G. Bunville & Y. Nozaki. The Reversible Transformations of β-Lactoglobulin at pH 7.5. *J. Am. Chem. Soc.*, 81:4032-4035, 1959.

D. Uhrín, J. Bramham, S. J. Winder & P. N. Barlow. Simultaneous CT-^{13}C and VT-^{15}N Chemical Shift Labelling: Application to 3D NOESY-CH_3NH and 3D ^{13}C, ^{15}N HSQC-NOESY-CH_3NH. *J. Biomol. NMR*, 18:253-259, 2000a.

D. Uhrin, S. Uhrinova, C. Leadbeater, J. Nairn, N. C. Price & P. N. Barlow. 3D $HCCH_3$-TOCSY for Resopnance Assignment of Methyl-Containing Side Chains in 13C-Labelled Proteins. *J. Magn. Reson.*, 142:288-293, 2000b.

S. Uhrínová, M. H. Smith, G. B. Jameson, D. Uhrín, L. Sawyer & P. N. Barlow. Structural Changed Accompanying the pH-Induced Dissociation of the β-lactoglobulin Dimer. *Biochemistry*, 39:3565-3574, 2000.

S. Uhrínová, D. Uhrín, H. Denton, M. Smith, L. Sawyer & P. N. Barlow. Complete Assignment of ^{1}H, ^{13}C and ^{15}N Chemical Shifts for Bovine β-lactoglobulin: Secondary Structure and Topology of the Native State is Retained in a Partially Unfolded Form. *J. Biomol. NMR*, 12:89-107, 1998.

H. Vis, R. Boelens, M. Mariani, R. Stroop, C. E. Vorgias, K. S. Wilson & R. Kaptein. ^{1}H, ^{13}C and ^{15}N Resonance Assignments and Secondary Structure Analysis of the HU Protein from Bacillus stearothermophilus Using 2D and 3D Double- and Triple-Resonance Heteronuclear Magnetic Resonance Spectroscopy, *Biochemistry*, 33: 14858 – 14870, 1994.

W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides & E. D. Laue. The CCPN Datamodel for NMR Spectroscopy: Development of a Software Pipeline. *Proteins*, 59:687-696, 2005.

G. Vriend. WHAT IF: A Molecular Modelling and Drug Design Program. *J. Mol. Graphics*, 8:52-56, 1990.

G. W. Vuister, R. Boelens, R. Kaptein, M. Burgering & P. C. M. van Zijl. Gradient-Enhanced 3D NOESY-HSQC Spectroscopy. *J. Biomol. NMR*, 2:301-305, 1992.

G. W. Vuister, G. M. Clore, A. M. Gronenborn, R. Powers, D. S. Garrett, R. Tschudin & A. Bax. Increased Resoltuion and Improved Spectral Quality in Four-Dimensional $^{13}C/^{13}C$-Separated HMQC-NOESY-HMQC Spectra Using Pulsed Field Gradients. *J. Magn. Reson. B*, 101:210-213, 1993.

J. Wang, T. Wang, E. R. P. Zuiderweg & G. M. Crippen. CASA: An Efficient Automated Assignment of Protein Mainchain NMR Data Using an Ordered Tree Search Algorithm. *J. Biomol. NMR*, 33:261-279, 2005.

S. S. Wijmenga & C. W. Hilbers. A 3D HMQC Homonuclear Hartmann-Hann NOESY Experiment. *J. Magn. Reson.*, 88:627-635, 1990.

D. S. Wishart & B. D. Sykes. The ^{13}C Chemical Shift Index: A Simple Method for the Identification of Protein Secondary Structure Using ^{13}C Chemical Shift Data. *J. Biomol. NMR* 4:171-180, 1994.

M. Wittekind & L. Mueller. A High Sensitivity 3D NMR Experiment to Correlate Amide Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *J. Magn. Reson. B*, 101:201-205, 1993.

K. Wuethrich. NMR of Proteins and Nucleic Acids. *Wiley-Interscience*, Chichester, 1986.

K. Wuethrich. Protein Structure Determination in Solution by Nuclear Magnetic Resonance Spectroscopy. *Science*, 243:45-50, 1989.

Y. Xia, D. Man & G. Zhu. 3D H_{aro}-Noesy-CH_3NH and C_{aro}-NOESY-CH_3NH Experiments for Double Labelled Proteins. *J. Biomol. NMR*, 19:355-360, 2001.

Y. Xia, A. Yee, C. H. Arrowsmith & X. Gao. $^1H^C$ and $^1H^N$ Total NOE Correlations in a Single 3D NMR Experiment. ^{15}N and ^{13}C Time-Sharing in t_1 and t_2 Dimensions for Simultaneous Data Acquisition. *J. Biomol. NMR*, 27:193-203, 2003.

T. Yamazaki, J. D. Forman-Kay & L. E. Kay. Two Dimensional NMR Experiments for Correlating $^{13}C\beta$ and $^1H\delta/\varepsilon$ Resonances of Aromatic Residues in ^{13}C-Labelled Proteins via Scalar Couplings. *J. Am. Chem. Soc.*, 115:11054-11055, 1993.

O. Zhang, L. E. Kay, J. P. Olivier & J. D. Forman-Kay. Backbone ^1H and ^{15}N Resonance Assignments of the N-terminal SH3 Domain of drk in Folded and Unfolded States Using Enhanced-Sensitivity PFG NMR Techniques. *J. Biomol. NMR*, 4:848-858, 1994.

O. Zhang & J. D. Forman-Kay. NMR Studies of Unfolded States of an SH3 Domain in Aqueous Solution and Denaturing Conditions. *Biochemistry*, 36:3959-3970, 1997.

W. Zhang & W. H. Gmeiner. A 3D NOESY-(HCACO)NH Experiment for the Measurement of NOEs Involving $^1H^\alpha$ in $^{13}C/^{15}N$-Labelled Proteins Dissolved in H_2O. *J. Biomol. NMR*, 8:357-359, 1996.

E. R. P. Zuiderweg, L. P. McIntosh, F. W. Dahlquist & S. W. Fesik. 3D ^{13}C-Resolved Proton NOESY of Uniformly ^{13}C-Labelled Proteins for the NMR Assignment and Structure Determination of Larger Molecules. *J. Magn. Reson.*, 86:210-216, 1990.

C. Zwahlen, S. J. F. Vincent, K. H. Gardner & L. E. Kay. Significantly Improved Resolution for NOE Correlations from Valine and Isoleucine ($C^{\gamma 2}$) Methyl Groups in ^{15}N, ^{13}C & ^{15}N, ^{13}C ^2H-labelled Proteins. *J. Am. Chem. Soc.*, 120:4825-4831, 1998.

Appendix A

BlgB Chemical Shift Assignments

No.	Residue	Atom	Shift
-2	VAL	HA	4.145
-2	VAL	HB	2.262
-2	VAL	HGA	1.064
-2	VAL	HGB	0.965
-2	VAL	CA	59.798
-2	VAL	CB	31.933
-2	VAL	CGA	20.719
-2	VAL	CGB	18.838
-1	PRO	HA	4.451
-1	PRO	HBA	1.814
-1	PRO	HBB	2.277
-1	PRO	HGA	1.975h
-1	PRO	HGB	1.991
-1	PRO	HDA	3.590
-1	PRO	HDB	3.734
-1	PRO	CA	63.313
-1	PRO	CB	32.148
-1	PRO	CG	27.451
-1	PRO	CD	50.905
0	MET	HA	4.369
0	MET	HBA	1.969
0	MET	HBB	1.969
0	MET	HGA	2.483
0	MET	HGB	2.532
0	MET	HN	8.318
0	MET	CA	55.626
0	MET	CB	33.577
0	MET	CG	31.910
0	MET	N	121.183
1	LEU	HA	4.335
1	LEU	HBA	1.525
1	LEU	HBB	1.525
1	LEU	HG	1.490
1	LEU	HDA	0.830
1	LEU	HDB	0.773
1	LEU	HN	8.104
1	LEU	CA	55.287
1	LEU	CB	42.578
1	LEU	CG	27.079
1	LEU	CDA	24.869
1	LEU	CDB	24.136
1	LEU	N	123.599
2	ILE	HA	4.053
2	ILE	HB	1.811
2	ILE	HGA	0.756
2	ILE	HG1A	1.075
2	ILE	HG1B	1.352
2	ILE	HD1	0.738
2	ILE	HN	8.034
2	ILE	CA	61.292
2	ILE	CB	38.444
2	ILE	CG1	27.359
2	ILE	CG2	17.749
2	ILE	CD1	12.912
2	ILE	N	121.683
3	VAL	HA	3.990
3	VAL	HB	1.951
3	VAL	HGA	0.825
3	VAL	HGB	0.825
3	VAL	HN	8.012
3	VAL	CA	62.764
3	VAL	CB	32.287
3	VAL	CGA	21.198
3	VAL	CGB	21.198
3	VAL	N	123.871
4	THR	HA	4.293
4	THR	HB	4.106
4	THR	HGA	1.041
4	THR	HN	7.921
4	THR	CA	61.298
4	THR	CB	69.893
4	THR	CG2	21.502
4	THR	N	115.866
5	GLN	HA	4.416
5	GLN	HBA	2.062
5	GLN	HBB	2.062
5	GLN	HGA	2.313
5	GLN	HGB	2.313
5	GLN	HN	8.166
5	GLN	HE1A	6.733
5	GLN	HE1B	7.375
5	GLN	CA	55.942
5	GLN	CB	29.253
5	GLN	CG	33.823
5	GLN	N	123.190
5	GLN	NE2	111.373
6	THR	HA	4.482
6	THR	HB	4.257
6	THR	HGA	0.981
6	THR	HN	7.639
6	THR	CA	60.545
6	THR	CB	70.576
6	THR	CG2	23.005
6	THR	N	111.167
7	MET	HA	4.216
7	MET	HBA	1.848
7	MET	HBB	1.639
7	MET	HGA	2.369
7	MET	HGB	2.539
7	MET	HN	7.671
7	MET	CA	56.636
7	MET	CB	33.220
7	MET	CG	32.487
7	MET	N	120.375
8	LYS	HA	4.212
8	LYS	HBA	1.798
8	LYS	HBB	1.798
8	LYS	HGA	1.418
8	LYS	HGB	1.522
8	LYS	HDA	1.686
8	LYS	HDB	1.686
8	LYS	HEA	3.006
8	LYS	HEB	3.006
8	LYS	HN	8.539
8	LYS	CA	57.127
8	LYS	CB	32.753
8	LYS	CG	25.083
8	LYS	CD	29.202
8	LYS	CE	42.267
8	LYS	N	127.225
9	GLY	HAA	3.817
9	GLY	HAB	3.919
9	GLY	HN	8.474
9	GLY	CA	45.855
9	GLY	N	109.254
10	LEU	HA	4.115
10	LEU	HBA	1.748
10	LEU	HBB	1.443
10	LEU	HG	1.437
10	LEU	HDA	1.006
10	LEU	HDB	0.953
10	LEU	HN	7.479
10	LEU	CA	56.227
10	LEU	CB	42.881
10	LEU	CG	27.253
10	LEU	CDA	26.911
10	LEU	CDB	23.878
10	LEU	N	120.031
11	ASP	HA	4.989
11	ASP	HBA	2.678
11	ASP	HBB	2.756
11	ASP	HN	8.555
11	ASP	CA	51.187
11	ASP	CB	38.299
11	ASP	N	126.586
12	ILE	HA	3.943
12	ILE	HB	2.036
12	ILE	HGA	1.032
12	ILE	HG1A	1.335
12	ILE	HG1B	1.410
12	ILE	HD1	0.870
12	ILE	HN	8.807

No.	Residue	Atom	Shift
12	ILE	CB	38.812
12	ILE	CG1	27.738
12	ILE	CG2	18.663
12	ILE	CD1	14.899
12	ILE	N	122.815
13	GLN	HA	3.871
13	GLN	HBA	2.067
13	GLN	HBB	1.963
13	GLN	HGA	2.442
13	GLN	HGB	2.442
13	GLN	HN	7.905
13	GLN	HE1A	6.792
13	GLN	HE1B	7.568
13	GLN	CA	58.324
13	GLN	CB	27.399
13	GLN	CG	33.662
13	GLN	N	117.157
13	GLN	NE2	112.351
14	LYS	HA	4.141
14	LYS	HBA	1.395
14	LYS	HBB	1.141
14	LYS	HGA	1.075
14	LYS	HGB	1.075
14	LYS	HDA	1.090
14	LYS	HDB	1.090
14	LYS	HEA	2.605
14	LYS	HEB	2.605
14	LYS	HN	7.155
14	LYS	CA	57.439
14	LYS	CB	33.398
14	LYS	CG	25.974
14	LYS	CD	29.347
14	LYS	CE	41.695
14	LYS	N	113.641
15	VAL	HA	4.752
15	VAL	HB	2.516
15	VAL	HGA	1.023
15	VAL	HGB	0.940
15	VAL	HN	7.336
15	VAL	CA	61.032
15	VAL	CB	31.269
15	VAL	CGA	22.427
15	VAL	CGB	19.466
15	VAL	N	110.225
16	ALA	HA	4.200
16	ALA	HB	1.480
16	ALA	HN	6.754
16	ALA	CA	52.675
16	ALA	CB	20.248
16	ALA	N	120.862
17	GLY	HAA	3.872
17	GLY	HAB	4.551
17	GLY	HN	9.125
17	GLY	CA	44.378
17	GLY	N	110.066
18	THR	HA	4.337
18	THR	HB	3.845
18	THR	HGA	1.030
18	THR	HN	8.329
18	THR	CA	64.247
18	THR	CB	69.906
18	THR	CG2	22.010
18	THR	N	118.314
19	TRP	HBA	2.162
19	TRP	HN	7.405
19	TRP	N	123.972
20	TYR	HA	5.126
20	TYR	HBA	2.743
20	TYR	HBB	2.743
20	TYR	HD1	7.018
20	TYR	HDA	7.018
20	TYR	HE1	6.692
20	TYR	HEA	6.692
20	TYR	HN	9.033
20	TYR	CA	56.687
20	TYR	CB	42.158
20	TYR	CD1	133.619
20	TYR	CD2	133.619
20	TYR	CE1	117.936
20	TYR	CE2	117.936
20	TYR	N	111.980
21	SER	HA	4.801
21	SER	HBA	3.843
21	SER	HBB	3.973
21	SER	HN	9.820
21	SER	CA	60.199
21	SER	CB	63.111
21	SER	N	119.467
22	LEU	HA	4.687
22	LEU	HBA	1.820
22	LEU	HBB	1.660
22	LEU	HG	1.729
22	LEU	HDA	1.160
22	LEU	HDB	1.082
22	LEU	HN	9.148
22	LEU	CA	56.811
22	LEU	CB	44.359
22	LEU	CG	27.787
22	LEU	CDA	24.234
22	LEU	CDB	24.234
22	LEU	N	131.338
23	ALA	HA	5.105
35	GLN	HBA	1.733
23	ALA	HB	1.173
23	ALA	HN	7.642
23	ALA	CA	50.859
23	ALA	CB	23.118
23	ALA	N	115.030
24	MET	HA	5.627
24	MET	HBA	2.365
24	MET	HBB	1.808
24	MET	HGA	2.412
24	MET	HGB	2.523
24	MET	HN	8.999
24	MET	CA	54.452
24	MET	CB	37.724
24	MET	CG	32.195
24	MET	N	115.210
25	ALA	HA	5.230
25	ALA	HB	1.254
25	ALA	HN	9.189
25	ALA	CA	51.485
25	ALA	CB	24.646
25	ALA	N	123.203
26	ALA	HA	5.281
26	ALA	HB	1.410
26	ALA	HN	8.362
26	ALA	CA	51.752
26	ALA	CB	23.549
26	ALA	N	119.834
27	SER	HA	4.791
27	SER	HBA	4.201
27	SER	HBB	3.859
27	SER	HN	8.548
27	SER	CA	60.151
27	SER	CB	64.275
27	SER	N	115.907
28	ASP	HA	5.069
28	ASP	HBA	2.837
28	ASP	HBB	2.837
28	ASP	HN	7.537
28	ASP	CA	51.653
28	ASP	CB	39.962
28	ASP	N	119.391
29	ILE	HA	3.603
29	ILE	HB	1.823
29	ILE	HGA	0.968
29	ILE	HG1A	1.111
29	ILE	HG1B	1.114
29	ILE	HD1	0.930
29	ILE	HN	8.807
29	ILE	CA	65.792
29	ILE	CB	38.605
29	ILE	CG1	28.870

No.	Residue	Atom	Shift	No.	Residue	Atom	Shift	No.	Residue	Atom	Shift
29	ILE	CG2	17.928	35	GLN	HBB	1.801	40	ARG	HGA	1.221
29	ILE	CD1	13.837	35	GLN	HGA	1.957	40	ARG	HGB	1.686
29	ILE	N	124.735	35	GLN	HGB	2.045	40	ARG	HDA	3.197
				35	GLN	HN	8.976	40	ARG	HDB	3.197
30	SER	HA	4.171	35	GLN	HE1A	6.439	40	ARG	HN	7.499
30	SER	HBA	3.863	35	GLN	HE1B	6.532	40	ARG	CA	54.653
30	SER	HBB	3.912	35	GLN	CA	59.074	40	ARG	CB	28.495
30	SER	HG	1.558	35	GLN	CB	28.262	40	ARG	CG	27.699
30	SER	HN	7.926	35	GLN	CG	32.775	40	ARG	CD	43.479
30	SER	CA	60.310	35	GLN	N	121.129	40	ARG	N	123.411
30	SER	CB	62.988	35	GLN	NE2	109.621				
30	SER	N	114.197					41	VAL	HA	4.606
				36	SER	HA	4.799	41	VAL	HB	1.801
31	LEU	HA	4.108	36	SER	HBA	3.898	41	VAL	HGA	0.650
31	LEU	HBA	1.723	36	SER	HBB	4.200	41	VAL	HGB	0.529
31	LEU	HBB	1.332	36	SER	HN	7.616	41	VAL	HN	6.848
31	LEU	HG	1.625	36	SER	CA	57.143	41	VAL	CA	57.421
31	LEU	HDA	0.948	36	SER	CB	63.915	41	VAL	CB	34.347
31	LEU	HDB	0.865	36	SER	N	107.549	41	VAL	CGA	22.029
31	LEU	HN	7.578					41	VAL	CGB	17.063
31	LEU	CA	56.845	37	ALA	HA	4.413	41	VAL	N	116.281
31	LEU	CB	41.698	37	ALA	HB	1.365				
31	LEU	CG	25.562	37	ALA	HN	7.449	42	TYR	HA	4.778
31	LEU	CDA	21.747	37	ALA	CA	51.642	42	TYR	HBA	2.719
31	LEU	CDB	21.747	37	ALA	CB	18.088	42	TYR	HBB	3.060
31	LEU	N	117.019	37	ALA	N	126.172	42	TYR	HD1	7.063
								42	TYR	HDA	7.063
32	LEU	HA	4.605	38	PRO	HA	4.183	42	TYR	HE1	6.770
32	LEU	HBA	1.613	38	PRO	HBA	1.882	42	TYR	HEA	6.770
32	LEU	HBB	1.223	38	PRO	HBB	2.358	42	TYR	HN	8.548
32	LEU	HG	1.352	38	PRO	HGA	2.002	42	TYR	CA	55.341
32	LEU	HDA	0.697	38	PRO	HGB	2.128	42	TYR	CB	38.761
32	LEU	HDB	0.697	38	PRO	HDA	3.781	42	TYR	CD1	133.228
32	LEU	HN	7.003	38	PRO	HDB	3.895	42	TYR	CD2	133.228
32	LEU	CA	54.184	38	PRO	CA	65.914	42	TYR	CE1	118.247
32	LEU	CB	48.520	38	PRO	CB	32.475	42	TYR	CE2	118.247
32	LEU	CG	26.317	38	PRO	CG	27.982	42	TYR	N	120.946
32	LEU	CDA	24.466	38	PRO	CD	50.305				
32	LEU	CDB	24.466					43	VAL	HA	3.715
32	LEU	N	111.353	39	LEU	HA	4.612	43	VAL	HB	1.869
				39	LEU	HBA	1.455	43	VAL	HGA	0.729
33	ASP	HA	4.218	39	LEU	HBB	1.455	43	VAL	HGB	0.620
33	ASP	HBA	2.763	39	LEU	HG	1.693	43	VAL	HN	8.564
33	ASP	HBB	3.077	39	LEU	HDA	0.914	43	VAL	CA	64.431
33	ASP	HN	7.700	39	LEU	HDB	0.914	43	VAL	CB	31.786
33	ASP	CA	56.108	39	LEU	HN	7.782	43	VAL	CGA	21.849
33	ASP	CB	38.350	39	LEU	CA	52.078	43	VAL	CGB	21.722
33	ASP	N	118.536	39	LEU	CB	40.498	43	VAL	N	125.260
				39	LEU	CG	27.427				
34	ALA	HA	4.516	39	LEU	CDA	26.667	44	GLU	HA	5.049
34	ALA	HB	1.399	39	LEU	CDB	23.222	44	GLU	HBA	2.028
34	ALA	HN	7.537	39	LEU	N	111.956	44	GLU	HBB	1.846
34	ALA	CA	51.501					44	GLU	HGA	2.268
34	ALA	CB	21.306	40	ARG	HA	3.911	44	GLU	HGB	2.354
34	ALA	N	120.513	40	ARG	HBA	1.328	44	GLU	HN	8.903
				40	ARG	HBB	2.199	44	GLU	CA	55.349
35	GLN	HA	4.035	49	THR	HA	5.016	44	GLU	CB	31.984

No.	Residue	Atom	Shift
44	GLU	CG	33.360
44	GLU	N	121.921
45	GLU	HA	5.203
45	GLU	HBA	2.181
45	GLU	HBB	2.265
45	GLU	HGA	1.876
45	GLU	HGB	2.023
45	GLU	HN	7.571
45	GLU	CA	54.775
45	GLU	CB	33.361
45	GLU	CG	33.138
45	GLU	N	116.231
46	LEU	HA	5.105
46	LEU	HBA	0.978
46	LEU	HBB	2.045
46	LEU	HG	1.326
46	LEU	HDA	0.154
46	LEU	HDB	0.485
46	LEU	HN	8.605
46	LEU	CA	53.573
46	LEU	CB	44.051
46	LEU	CG	27.324
46	LEU	CDA	24.423
46	LEU	CDB	25.064
46	LEU	N	122.031
47	LYS	HA	5.054
47	LYS	HBA	1.581
47	LYS	HBB	1.801
47	LYS	HGA	1.278
47	LYS	HGB	1.278
47	LYS	HDA	1.553
47	LYS	HDB	1.553
47	LYS	HEA	2.868
47	LYS	HEB	2.868
47	LYS	HN	9.095
47	LYS	CA	52.975
47	LYS	CB	35.144
47	LYS	CG	24.454
47	LYS	CD	29.218
47	LYS	CE	41.989
47	LYS	N	122.356
48	PRO	HA	5.124
48	PRO	HBA	1.958
48	PRO	HBB	2.004
48	PRO	HGA	1.889
48	PRO	HGB	2.153
48	PRO	HDA	3.954
48	PRO	HDB	3.331
48	PRO	CA	62.169
48	PRO	CB	32.482
48	PRO	CG	27.782
48	PRO	CD	50.884
49	THR	HB	4.761
49	THR	HGA	1.119
49	THR	HN	8.537
49	THR	CA	58.906
49	THR	CB	68.778
49	THR	CG2	20.906
49	THR	N	114.207
50	PRO	HA	4.327
50	PRO	HBA	1.848
50	PRO	HBB	2.425
50	PRO	HGA	2.001
50	PRO	HGB	2.132
50	PRO	HDA	3.824
50	PRO	HDB	3.951
50	PRO	CA	65.443
50	PRO	CB	31.869
50	PRO	CG	27.833
50	PRO	CD	51.140
51	GLU	HA	4.366
51	GLU	HBA	1.904
51	GLU	HBB	2.223
51	GLU	HGA	2.404
51	GLU	HGB	2.535
51	GLU	HN	7.421
51	GLU	CA	56.214
51	GLU	CB	28.056
51	GLU	CG	33.985
51	GLU	N	112.155
52	GLY	HAA	3.562
52	GLY	HAB	4.242
52	GLY	HN	8.225
52	GLY	CA	45.623
52	GLY	N	106.918
53	ASP	HA	5.079
53	ASP	HBA	2.687
53	ASP	HBB	3.065
53	ASP	HN	7.115
53	ASP	CA	52.964
53	ASP	CB	39.449
53	ASP	N	115.015
54	LEU	HA	4.612
54	LEU	HBA	0.603
54	LEU	HBB	1.076
54	LEU	HG	0.878
54	LEU	HDA	-0.006
54	LEU	HDB	0.271
54	LEU	HN	8.742
54	LEU	CA	54.020
54	LEU	CB	46.124
54	LEU	CG	26.006
54	LEU	CDA	23.877
63	ASN	ND2	112.904
54	LEU	CDB	23.031
54	LEU	N	119.305
55	GLU	HA	4.815
55	GLU	HBA	1.931
55	GLU	HBB	1.931
55	GLU	HGA	2.158
55	GLU	HGB	2.158
55	GLU	HN	8.607
55	GLU	CA	55.034
55	GLU	CB	29.219
55	GLU	CG	33.366
55	GLU	N	125.926
56	ILE	HA	4.309
56	ILE	HB	1.592
56	ILE	HGA	0.508
56	ILE	HG1A	1.292
56	ILE	HG1B	1.292
56	ILE	HD1	0.258
56	ILE	HN	9.203
56	ILE	CA	60.772
56	ILE	CB	39.768
56	ILE	CG1	26.267
56	ILE	CG2	17.059
56	ILE	CD1	12.800
56	ILE	N	126.413
57	LEU	HA	5.083
57	LEU	HBA	1.545
57	LEU	HBB	1.757
57	LEU	HG	1.525
57	LEU	HDA	0.774
57	LEU	HDB	0.774
57	LEU	HN	8.435
57	LEU	CA	54.731
57	LEU	CB	42.587
57	LEU	CG	28.579
57	LEU	CDA	24.892
57	LEU	CDB	24.152
57	LEU	N	128.617
58	LEU	HA	5.364
58	LEU	HBA	1.465
58	LEU	HBB	1.342
58	LEU	HG	1.584
58	LEU	HDA	0.589
58	LEU	HDB	0.541
58	LEU	HN	9.015
58	LEU	CA	53.265
58	LEU	CB	46.167
58	LEU	CG	25.626
58	LEU	CDA	25.630
58	LEU	CDB	26.175
58	LEU	N	123.291
59	GLN	HA	5.427

No.	Residue	Atom	Shift
59	GLN	HBA	2.121
59	GLN	HBB	2.237
59	GLN	HGA	2.338
59	GLN	HGB	2.260
59	GLN	HN	9.150
59	GLN	HE1A	7.137
59	GLN	HE1B	7.202
59	GLN	CA	53.551
59	GLN	CB	31.934
59	GLN	CG	33.960
59	GLN	N	117.709
59	GLN	NE2	111.793
60	LYS	HA	4.630
60	LYS	HBA	1.554
60	LYS	HBB	1.554
60	LYS	HGA	1.356
60	LYS	HGB	1.356
60	LYS	HDA	1.499
60	LYS	HDB	1.499
60	LYS	HEA	2.941
60	LYS	HEB	2.941
60	LYS	HN	8.892
60	LYS	CA	55.346
60	LYS	CB	37.266
60	LYS	CG	24.586
60	LYS	CD	29.229
60	LYS	CE	42.419
60	LYS	N	121.331
61	TRP	HA	5.057
61	TRP	HBA	2.969
61	TRP	HBB	3.314
61	TRP	HD1	7.328
61	TRP	HN	9.679
61	TRP	N	129.765
62	GLU	HA	4.435
62	GLU	HBA	1.902
62	GLU	HBB	1.710
62	GLU	HGA	2.203
62	GLU	HGB	2.203
62	GLU	HN	8.792
62	GLU	CA	54.768
62	GLU	CB	30.731
62	GLU	CG	32.764
62	GLU	N	126.961
63	ASN	HA	4.155
63	ASN	HBA	2.569
63	ASN	HBB	2.889
63	ASN	HN	8.950
63	ASN	HD1A	6.783
63	ASN	HD1B	7.464
63	ASN	CA	54.082
63	ASN	CB	37.467
63	ASN	N	122.260
64	GLY	HAA	3.410
64	GLY	HAB	3.936
64	GLY	HN	7.013
64	GLY	CA	45.860
64	GLY	N	101.636
65	GLU	HA	4.093
65	GLU	HBA	1.850
65	GLU	HBB	1.693
65	GLU	HGA	2.126
65	GLU	HGB	2.225
65	GLU	HN	6.798
65	GLU	CA	53.232
65	GLU	CB	31.104
65	GLU	CG	31.875
65	GLU	N	116.489
66	CYS	HA	4.858
66	CYS	HBA	2.861
66	CYS	HBB	3.090
66	CYS	HN	8.624
66	CYS	CA	54.716
66	CYS	CB	38.970
66	CYS	N	120.776
67	ALA	HA	4.533
67	ALA	HB	1.333
67	ALA	HN	8.906
67	ALA	CA	51.503
67	ALA	CB	21.384
67	ALA	N	133.502
68	GLN	HA	5.106
68	GLN	HBA	2.047
68	GLN	HBB	1.936
68	GLN	HGA	2.410
68	GLN	HGB	2.237
68	GLN	HN	8.310
68	GLN	HE1A	7.103
68	GLN	HE1B	7.678
68	GLN	CA	54.729
68	GLN	CB	31.001
68	GLN	CG	34.259
68	GLN	N	118.502
68	GLN	NE2	113.358
69	LYS	HA	4.640
69	LYS	HBA	1.635
69	LYS	HBB	1.446
69	LYS	HGA	1.216
69	LYS	HGB	1.216
69	LYS	HDA	1.554
69	LYS	HDB	1.554
69	LYS	HEA	2.835
69	LYS	HEB	2.739
78	ILE	HA	4.502
69	LYS	HN	9.105
69	LYS	CA	54.412
69	LYS	CB	36.044
69	LYS	CG	24.465
69	LYS	CD	28.630
69	LYS	CE	41.874
69	LYS	N	123.755
70	LYS	HA	5.152
70	LYS	HBA	1.635
70	LYS	HBB	1.635
70	LYS	HGA	1.226
70	LYS	HGB	1.090
70	LYS	HDA	1.544
70	LYS	HDB	1.544
70	LYS	HEA	2.810
70	LYS	HEB	2.810
70	LYS	HN	8.475
70	LYS	CA	55.077
70	LYS	CB	34.748
70	LYS	CG	24.775
70	LYS	CD	29.513
70	LYS	CE	41.880
70	LYS	N	123.587
71	ILE	HA	4.397
71	ILE	HB	1.604
71	ILE	HGA	0.658
71	ILE	HG1A	0.913
71	ILE	HG1B	1.256
71	ILE	HD1	0.543
71	ILE	HN	8.907
71	ILE	CA	59.615
71	ILE	CB	41.653
71	ILE	CG1	27.072
71	ILE	CG2	17.247
71	ILE	CD1	13.665
71	ILE	N	123.982
72	ILE	HA	4.372
72	ILE	HB	1.844
72	ILE	HGA	0.715
72	ILE	HG1A	1.138
72	ILE	HG1B	1.383
72	ILE	HD1	0.727
72	ILE	HN	8.435
72	ILE	CA	59.836
72	ILE	CB	37.525
72	ILE	CG1	27.570
72	ILE	CG2	17.356
72	ILE	CD1	11.639
72	ILE	N	125.401
73	ALA	HA	5.012
73	ALA	HB	0.950
73	ALA	HN	9.205
73	ALA	CA	49.711

No.	Residue	Atom	Shift	No.	Residue	Atom	Shift	No.	Residue	Atom	Shift
73	ALA	CB	19.436	78	ILE	HB	1.625	83	LYS	HA	4.968
73	ALA	N	133.255	78	ILE	HGA	0.874	83	LYS	HBA	1.743
				78	ILE	HG1A	1.086	83	LYS	HBB	1.769
74	GLU	HA	4.555	78	ILE	HG1B	1.381	83	LYS	HGA	1.529
74	GLU	HBA	2.145	78	ILE	HD1	0.764	83	LYS	HGB	1.330
74	GLU	HBB	1.986	78	ILE	HN	7.926	83	LYS	HDA	1.691
74	GLU	HGA	2.466	78	ILE	CA	57.429	83	LYS	HDB	1.711
74	GLU	HGB	2.466	78	ILE	CB	39.035	83	LYS	HEA	2.963
74	GLU	HN	9.377	78	ILE	CG1	27.149	83	LYS	HEB	2.963
74	GLU	CA	55.035	78	ILE	CG2	17.562	83	LYS	HN	9.421
74	GLU	CB	30.194	78	ILE	CD1	12.173	83	LYS	CA	55.638
74	GLU	CG	33.388	78	ILE	N	122.136	83	LYS	CB	34.849
74	GLU	N	124.487					83	LYS	CG	24.872
				79	PRO	HA	4.203	83	LYS	CD	29.528
75	LYS	HA	4.067	79	PRO	HBA	1.937	83	LYS	CE	41.987
75	LYS	HBA	1.876	79	PRO	HBB	2.419	83	LYS	N	123.086
75	LYS	HBB	1.960	79	PRO	HGA	2.034				
75	LYS	HGA	1.489	79	PRO	HGB	2.149	84	ILE	HA	4.834
75	LYS	HGB	1.241	79	PRO	HDA	3.722	84	ILE	HB	1.696
75	LYS	HDA	1.666	79	PRO	HDB	4.010	84	ILE	HGA	0.733
75	LYS	HDB	1.666	79	PRO	CA	64.543	84	ILE	HG1A	1.175
75	LYS	HEA	2.977	79	PRO	CB	32.477	84	ILE	HG1B	1.175
75	LYS	HEB	2.977	79	PRO	CG	27.730	84	ILE	HD1	0.608
75	LYS	HN	8.395	79	PRO	CD	51.464	84	ILE	HN	7.650
75	LYS	CA	57.146					84	ILE	CA	59.489
75	LYS	CB	32.573	80	ALA	HA	4.047	84	ILE	CB	43.778
75	LYS	CG	23.849	80	ALA	HB	1.763	84	ILE	CG1	24.467
75	LYS	CD	29.820	80	ALA	HN	7.788	84	ILE	CG2	19.132
75	LYS	CE	42.394	80	ALA	CA	52.712	84	ILE	CD1	14.279
75	LYS	N	123.032	80	ALA	CB	19.893	84	ILE	N	115.772
				80	ALA	N	114.991				
76	THR	HA	4.778					85	ASP	HA	4.994
76	THR	HB	4.429	81	VAL	HA	4.913	85	ASP	HBA	2.879
76	THR	HG1	4.648	81	VAL	HB	1.709	85	ASP	HBB	2.703
76	THR	HGA	1.186	81	VAL	HGA	0.841	85	ASP	HN	8.738
76	THR	HN	8.368	81	VAL	HGB	0.805	85	ASP	CA	52.390
76	THR	CA	60.131	81	VAL	HN	7.256	85	ASP	CB	38.158
76	THR	CB	72.287	81	VAL	CA	61.877	85	ASP	N	119.081
76	THR	CG2	20.824	81	VAL	CB	33.486				
76	THR	N	112.498	81	VAL	CGA	22.131	86	ALA	HA	4.609
				81	VAL	CGB	21.234	86	ALA	HB	1.277
77	LYS	HA	4.119	81	VAL	N	117.391	86	ALA	HN	7.155
77	LYS	HBA	1.797					86	ALA	CA	51.787
77	LYS	HBB	1.903	82	PHE	HA	5.182	86	ALA	CB	22.760
77	LYS	HGA	1.460	82	PHE	HBA	2.443	86	ALA	N	121.793
77	LYS	HGB	1.460	82	PHE	HBB	2.718				
77	LYS	HDA	1.662	82	PHE	HD1	6.794	87	LEU	HA	3.839
77	LYS	HDB	1.662	82	PHE	HDA	6.794	87	LEU	HBA	1.876
77	LYS	HEA	2.973	82	PHE	HE1	7.134	87	LEU	HBB	1.573
77	LYS	HEB	2.973	82	PHE	HEA	7.134	87	LEU	HG	1.273
77	LYS	HN	8.215	82	PHE	HN	9.498	87	LEU	HDA	0.773
77	LYS	CA	57.673	82	PHE	CA	55.885	87	LEU	HDB	0.687
77	LYS	CB	32.479	82	PHE	CB	43.793	87	LEU	HN	8.970
77	LYS	CG	25.091	82	PHE	CD1	131.350	87	LEU	CA	55.674
77	LYS	CD	29.232	82	PHE	CD2	131.350	87	LEU	CB	38.414
77	LYS	CE	42.279	82	PHE	N	126.260	87	LEU	CG	26.463
77	LYS	N	116.433	92	VAL	HGB	0.629	87	LEU	CDA	22.499

No.	Residue	Atom	Shift
87	LEU	CDB	20.008
87	LEU	N	114.362
88	ASN	HA	4.196
88	ASN	HBA	2.807
88	ASN	HBB	3.145
88	ASN	HN	8.635
88	ASN	HD1A	6.736
88	ASN	HD1B	7.448
88	ASN	CA	54.935
88	ASN	CB	37.827
88	ASN	N	111.561
88	ASN	ND2	112.203
89	GLU	HA	4.932
89	GLU	HBA	1.987
89	GLU	HBB	1.987
89	GLU	HGA	2.669
89	GLU	HGB	2.308
89	GLU	HN	8.192
89	GLU	CA	53.866
89	GLU	CB	32.477
89	GLU	CG	30.113
89	GLU	N	117.030
90	ASN	HA	5.720
90	ASN	HBA	3.384
90	ASN	HBB	2.766
90	ASN	HN	9.014
90	ASN	HD1A	6.839
90	ASN	HD1B	7.424
90	ASN	CA	52.822
90	ASN	CB	39.953
90	ASN	N	115.480
90	ASN	ND2	109.318
91	LYS	HA	5.543
91	LYS	HBA	1.638
91	LYS	HBB	1.638
91	LYS	HGA	1.437
91	LYS	HGB	1.437
91	LYS	HDA	1.470
91	LYS	HDB	1.615
91	LYS	HEA	2.832
91	LYS	HEB	2.832
91	LYS	HN	8.504
91	LYS	HZ	8.492
91	LYS	CA	55.638
91	LYS	CB	37.850
91	LYS	CG	25.175
91	LYS	CD	29.416
91	LYS	CE	42.397
91	LYS	N	121.509
92	VAL	HA	4.806
92	VAL	HB	1.748
92	VAL	HGA	0.754
92	VAL	HN	8.995
92	VAL	CA	61.351
92	VAL	CB	35.152
92	VAL	CGA	22.677
92	VAL	CGB	20.398
92	VAL	N	121.169
93	LEU	HA	5.342
93	LEU	HBA	1.387
93	LEU	HBB	1.387
93	LEU	HG	1.549
93	LEU	HDA	0.752
93	LEU	HDB	0.704
93	LEU	HN	9.355
93	LEU	CA	53.818
93	LEU	CB	44.372
93	LEU	CG	27.958
93	LEU	CDA	26.418
93	LEU	CDB	26.030
93	LEU	N	125.831
94	VAL	HA	4.272
94	VAL	HB	2.240
94	VAL	HGA	0.800
94	VAL	HGB	0.754
94	VAL	HN	9.279
94	VAL	CA	63.402
94	VAL	CB	31.529
94	VAL	CGA	20.839
94	VAL	CGB	20.167
94	VAL	N	123.993
95	LEU	HA	4.244
95	LEU	HBA	1.740
95	LEU	HBB	1.403
95	LEU	HG	1.660
95	LEU	HDA	0.787
95	LEU	HDB	0.676
95	LEU	HN	8.379
95	LEU	CA	57.452
95	LEU	CB	42.609
95	LEU	CG	28.671
95	LEU	CDA	25.355
95	LEU	CDB	23.169
95	LEU	N	127.862
96	ASP	HA	4.919
96	ASP	HBA	2.316
96	ASP	HBB	3.133
96	ASP	HN	7.133
96	ASP	CA	54.189
96	ASP	CB	43.195
96	ASP	N	109.684
97	THR	HA	4.115
97	THR	HB	4.125
105	PHE	HEA	6.994
97	THR	HG1	2.109
97	THR	HGA	0.213
97	THR	HN	7.749
97	THR	CA	60.402
97	THR	CB	69.470
97	THR	CG2	17.490
97	THR	N	118.859
98	ASP	HA	4.629
98	ASP	HBA	2.518
98	ASP	HBB	3.235
98	ASP	HN	6.712
98	ASP	CA	52.388
98	ASP	CB	41.380
98	ASP	N	122.840
99	TYR	HA	3.735
99	TYR	HBA	3.048
99	TYR	HBB	3.580
99	TYR	HD1	6.804
99	TYR	HDA	6.804
99	TYR	HE1	6.815
99	TYR	HEA	6.815
99	TYR	HN	8.974
99	TYR	CA	65.229
99	TYR	CB	36.604
99	TYR	CD1	132.941
99	TYR	CD2	132.941
99	TYR	CE1	118.380
99	TYR	CE2	118.380
99	TYR	N	117.658
100	LYS	HA	4.274
100	LYS	HBA	1.916
100	LYS	HBB	1.767
100	LYS	HGA	1.496
100	LYS	HGB	1.496
100	LYS	HDA	1.733
100	LYS	HDB	1.733
100	LYS	HEA	3.026
100	LYS	HEB	3.026
100	LYS	HN	8.944
100	LYS	CA	57.098
100	LYS	CB	34.566
100	LYS	CG	25.116
100	LYS	CD	29.182
100	LYS	CE	42.276
100	LYS	N	119.332
101	LYS	HA	4.749
101	LYS	HBA	2.004
101	LYS	HBB	2.739
101	LYS	HGA	1.391
101	LYS	HGB	1.479
101	LYS	HDA	1.798
101	LYS	HDB	1.798
101	LYS	HEA	3.027

No.	Residue	Atom	Shift	No.	Residue	Atom	Shift	No.	Residue	Atom	Shift
101	LYS	HEB	3.027	105	PHE	HN	8.829	110	SER	CA	61.267
101	LYS	HN	9.000	105	PHE	CA	55.962	110	SER	CB	63.115
101	LYS	CA	59.200	105	PHE	CB	42.376	110	SER	N	122.098
101	LYS	CB	35.776	105	PHE	CD1	131.354				
101	LYS	CG	25.356	105	PHE	CD2	131.354	111	ALA	HA	4.312
101	LYS	CD	29.059	105	PHE	CE1	130.503	111	ALA	HB	1.365
101	LYS	CE	42.292	105	PHE	CE2	130.503	111	ALA	HN	8.233
101	LYS	N	118.343	105	PHE	N	120.949	111	ALA	CA	53.632
								111	ALA	CB	19.221
102	TYR	HA	6.520	106	CYS	HA	5.513	111	ALA	N	123.189
102	TYR	HBA	2.756	106	CYS	HBA	2.600				
102	TYR	HBB	3.075	106	CYS	HBB	2.990	112	GLU	HA	4.793
102	TYR	HD1	6.414	106	CYS	HG	7.076	112	GLU	HBA	1.946
102	TYR	HDA	6.414	106	CYS	HN	9.595	112	GLU	HBB	2.168
102	TYR	HE1	6.570	106	CYS	CA	55.970	112	GLU	HGA	2.449
102	TYR	HEA	6.570	106	CYS	CB	49.396	112	GLU	HGB	2.272
102	TYR	HN	8.788	106	CYS	N	116.881	112	GLU	HN	7.682
102	TYR	CA	56.762					112	GLU	CA	53.261
102	TYR	CB	44.077	107	MET	HA	5.676	112	GLU	CB	28.031
102	TYR	CE1	117.224	107	MET	HBA	2.033	112	GLU	CG	32.184
102	TYR	CE2	117.224	107	MET	HBB	2.033	112	GLU	N	112.900
102	TYR	N	115.957	107	MET	HGA	2.650				
				107	MET	HGB	2.733	113	PRO	HA	4.111
103	LEU	HA	4.357	107	MET	HN	9.777	113	PRO	HBA	1.875
103	LEU	HBA	0.973	107	MET	CA	54.320	113	PRO	HBB	2.141
103	LEU	HBB	0.325	107	MET	CB	37.231	113	PRO	HGA	1.974
103	LEU	HG	1.238	107	MET	CG	32.501	113	PRO	HGB	2.077
103	LEU	HDA	0.365	107	MET	N	120.775	113	PRO	HDA	3.514
103	LEU	HDB	0.239					113	PRO	HDB	3.675
103	LEU	HN	9.112	108	GLU	HA	4.724	113	PRO	CA	66.340
103	LEU	CA	56.855	108	GLU	HBA	2.066	113	PRO	CB	32.158
103	LEU	CB	44.102	108	GLU	HBB	2.066	113	PRO	CG	27.748
103	LEU	CG	27.714	108	GLU	HGA	1.724	113	PRO	CD	50.024
103	LEU	CDA	25.999	108	GLU	HGB	1.920				
103	LEU	CDB	26.272	108	GLU	HN	8.377	114	GLU	HA	4.002
103	LEU	N	124.936	108	GLU	CA	55.044	114	GLU	HBA	2.111
				108	GLU	CB	32.471	114	GLU	HBB	2.111
104	LEU	HA	5.708	108	GLU	CG	32.689	114	GLU	HGA	2.492
104	LEU	HBA	2.182	108	GLU	N	113.821	114	GLU	HGB	2.492
104	LEU	HBB	1.466					114	GLU	HN	8.300
104	LEU	HG	1.999	109	ASN	HA	5.119	114	GLU	CA	59.223
104	LEU	HDA	1.093	109	ASN	HBA	2.914	114	GLU	CB	27.759
104	LEU	HDB	0.978	109	ASN	HBB	3.052	114	GLU	CG	33.980
104	LEU	HN	9.288	109	ASN	HN	9.275	114	GLU	N	113.805
104	LEU	CA	52.693	109	ASN	HD1A	6.612				
104	LEU	CB	44.066	109	ASN	HD1B	7.357	115	GLN	HA	4.380
104	LEU	CG	27.402	109	ASN	CA	51.965	115	GLN	HBA	2.217
104	LEU	CDA	24.440	109	ASN	CB	39.034	115	GLN	HBB	1.955
104	LEU	CDB	26.072	109	ASN	N	120.199	115	GLN	HGA	2.379
104	LEU	N	124.836	109	ASN	ND2	111.178	115	GLN	HGB	2.467
								115	GLN	HN	7.671
105	PHE	HA	5.772	110	SER	HA	3.882	115	GLN	HE1B	6.843
105	PHE	HBA	2.726	110	SER	HBA	3.693	115	GLN	CA	57.101
105	PHE	HBB	3.169	110	SER	HBB	4.081	115	GLN	CB	29.004
105	PHE	HD1	6.847	110	SER	HG	4.919	115	GLN	CG	34.257
105	PHE	HDA	6.847	110	SER	HN	9.775	115	GLN	N	115.830
105	PHE	HE1	6.994	122	LEU	HA	5.898	115	GLN	NE2	112.890

No.	Residue	Atom	Shift
116	SER	HA	4.580
116	SER	HBA	3.732
116	SER	HBB	4.700
116	SER	HN	7.509
116	SER	CA	58.533
116	SER	CB	65.342
116	SER	N	109.550
117	LEU	HA	4.885
117	LEU	HBA	1.767
117	LEU	HBB	1.767
117	LEU	HG	1.480
117	LEU	HDA	0.866
117	LEU	HDB	0.809
117	LEU	HN	7.182
117	LEU	CA	56.529
117	LEU	CB	42.858
117	LEU	CDA	23.761
117	LEU	CDB	26.914
117	LEU	N	126.595
118	ALA	HA	5.338
118	ALA	HB	1.386
118	ALA	HN	8.708
118	ALA	CA	51.482
118	ALA	CB	22.476
118	ALA	N	128.238
119	CYS	HA	5.567
119	CYS	HBA	2.660
119	CYS	HBB	3.019
119	CYS	HN	9.446
119	CYS	CA	55.650
119	CYS	CB	50.571
119	CYS	N	118.583
120	GLN	HA	5.061
120	GLN	HBA	1.939
120	GLN	HBB	2.164
120	GLN	HGA	1.227
120	GLN	HGB	2.315
120	GLN	HN	9.191
120	GLN	HE1A	7.140
120	GLN	HE1B	7.702
120	GLN	CA	55.606
120	GLN	CB	32.753
120	GLN	N	118.487
120	GLN	NE2	108.570
121	CYS	HA	4.576
121	CYS	HBA	2.143
121	CYS	HBB	1.186
121	CYS	HN	7.818
121	CYS	CA	56.819
121	CYS	CB	27.546
121	CYS	N	119.979
122	LEU	HBA	1.859
122	LEU	HBB	1.646
122	LEU	HG	1.686
122	LEU	HDA	1.006
122	LEU	HDB	0.337
122	LEU	HN	9.668
122	LEU	CA	52.674
122	LEU	CB	44.755
122	LEU	CG	26.648
122	LEU	CDA	24.607
122	LEU	CDB	25.257
122	LEU	N	128.443
123	VAL	HA	6.114
123	VAL	HB	2.801
123	VAL	HGA	1.162
123	VAL	HGB	1.234
123	VAL	HN	9.394
123	VAL	CA	58.926
123	VAL	CB	37.068
123	VAL	CGA	24.546
123	VAL	CGB	19.120
123	VAL	N	111.370
124	ARG	HA	3.869
124	ARG	HBA	1.692
124	ARG	HBB	1.774
124	ARG	HGA	0.949
124	ARG	HGB	0.720
124	ARG	HDA	1.613
124	ARG	HDB	2.399
124	ARG	HN	7.881
124	ARG	CA	56.843
124	ARG	CB	33.392
124	ARG	CG	25.919
124	ARG	CD	42.850
124	ARG	N	119.542
125	THR	HA	4.807
125	THR	HB	4.205
125	THR	HGA	1.137
125	THR	HN	7.445
125	THR	CA	57.293
125	THR	CB	69.945
125	THR	CG2	21.769
125	THR	N	108.311
126	PRO	HA	4.018
126	PRO	HBA	0.831
126	PRO	HBB	1.811
126	PRO	HGA	1.481
126	PRO	HGB	1.729
126	PRO	HDA	3.114
126	PRO	HDB	3.520
126	PRO	CA	62.470
126	PRO	CB	28.348
136	PHE	CE2	131.418
126	PRO	CG	26.902
126	PRO	CD	49.711
127	GLU	HA	4.311
127	GLU	HBA	1.741
127	GLU	HBB	1.990
127	GLU	HGA	2.264
127	GLU	HGB	2.264
127	GLU	HN	8.224
127	GLU	CA	54.789
127	GLU	CB	30.678
127	GLU	CG	32.773
127	GLU	N	119.829
128	VAL	HA	3.900
128	VAL	HB	2.176
128	VAL	HGA	1.096
128	VAL	HGB	0.930
128	VAL	HN	8.225
128	VAL	CA	63.223
128	VAL	CB	31.172
128	VAL	CGA	20.757
128	VAL	CGB	21.458
128	VAL	N	118.524
129	ASP	HA	4.766
129	ASP	HBA	2.778
129	ASP	HBB	2.865
129	ASP	HN	6.511
129	ASP	CA	53.019
129	ASP	CB	39.932
129	ASP	N	123.755
130	ASP	HA	4.324
130	ASP	HBA	2.816
130	ASP	HBB	2.816
130	ASP	HN	9.033
130	ASP	CA	56.931
130	ASP	CB	37.859
130	ASP	N	123.583
131	GLU	HA	3.992
131	GLU	HBA	2.154
131	GLU	HBB	2.012
131	GLU	HGA	2.425
131	GLU	HGB	2.425
131	GLU	HN	8.299
131	GLU	CA	58.935
131	GLU	CB	27.470
131	GLU	CG	33.100
131	GLU	N	120.962
132	ALA	HA	3.222
132	ALA	HB	1.166
132	ALA	HN	7.533
132	ALA	CA	54.941
132	ALA	CB	19.312

No.	Residue	Atom	Shift
132	ALA	N	121.187
133	LEU	HA	3.983
133	LEU	HBA	1.501
133	LEU	HBB	1.835
133	LEU	HG	1.741
133	LEU	HDA	0.888
133	LEU	HDB	0.809
133	LEU	HN	7.704
133	LEU	CA	57.718
133	LEU	CB	41.408
133	LEU	CG	26.744
133	LEU	CDA	25.418
133	LEU	CDB	23.329
133	LEU	N	116.354
134	GLU	HA	4.070
134	GLU	HBA	2.123
134	GLU	HBB	2.123
134	GLU	HGA	2.422
134	GLU	HGB	2.562
134	GLU	HN	7.966
134	GLU	CA	59.213
134	GLU	CB	27.548
134	GLU	CG	33.346
134	GLU	N	119.650
135	LYS	HA	3.837
135	LYS	HBA	1.619
135	LYS	HBB	1.461
135	LYS	HGA	1.112
135	LYS	HGB	1.278
135	LYS	HDA	1.425
135	LYS	HDB	1.134
135	LYS	HEA	2.933
135	LYS	HEB	2.933
135	LYS	HN	7.875
135	LYS	CA	59.017
135	LYS	CB	31.964
135	LYS	CG	25.078
135	LYS	CD	28.658
135	LYS	CE	42.318
135	LYS	N	119.602
136	PHE	HA	4.258
136	PHE	HBA	3.132
136	PHE	HBB	3.299
136	PHE	HD1	7.203
136	PHE	HDA	7.203
136	PHE	HE1	7.296
136	PHE	HEA	7.296
136	PHE	HN	8.486
136	PHE	CA	60.100
136	PHE	CB	39.409
136	PHE	CD1	131.910
136	PHE	CD2	131.910
136	PHE	CE1	131.418
136	PHE	N	121.489
137	ASP	HA	3.970
137	ASP	HBA	2.719
137	ASP	HBB	2.881
137	ASP	HN	8.680
137	ASP	CA	56.337
137	ASP	CB	37.853
137	ASP	N	117.022
138	LYS	HA	3.885
138	LYS	HBA	1.811
138	LYS	HBB	1.830
138	LYS	HGA	1.368
138	LYS	HGB	1.491
138	LYS	HDA	1.612
138	LYS	HDB	1.612
138	LYS	HEA	2.922
138	LYS	HEB	2.922
138	LYS	HN	7.879
138	LYS	CA	59.515
138	LYS	CB	32.433
138	LYS	CG	25.335
138	LYS	CD	29.220
138	LYS	CE	42.284
138	LYS	N	119.605
139	ALA	HA	4.045
139	ALA	HB	1.332
139	ALA	HN	7.771
139	ALA	CA	54.380
139	ALA	CB	17.696
139	ALA	N	121.588
140	LEU	HA	3.862
140	LEU	HBA	1.334
140	LEU	HBB	1.239
140	LEU	HG	1.270
140	LEU	HDA	0.107
140	LEU	HDB	-0.079
140	LEU	HN	7.415
140	LEU	CA	55.312
140	LEU	CB	41.433
140	LEU	CG	26.495
140	LEU	CDA	23.121
140	LEU	CDB	24.115
140	LEU	N	114.138
141	LYS	HA	3.875
141	LYS	HBA	1.778
141	LYS	HBB	1.875
141	LYS	HGA	1.349
141	LYS	HGB	1.349
141	LYS	HDA	1.633
141	LYS	HDB	1.475
141	LYS	HEA	2.912
151	PHE	HA	5.033
141	LYS	HEB	2.912
141	LYS	HN	7.102
141	LYS	HZ	7.561
141	LYS	CA	59.486
141	LYS	CB	32.497
141	LYS	CG	24.566
141	LYS	CD	29.221
141	LYS	CE	41.536
141	LYS	N	119.668
142	ALA	HA	4.393
142	ALA	HB	1.336
142	ALA	HN	7.811
142	ALA	CA	51.672
142	ALA	CB	19.251
142	ALA	N	118.346
143	LEU	HA	4.680
143	LEU	HBA	1.636
143	LEU	HBB	1.114
143	LEU	HG	1.628
143	LEU	HDA	0.769
143	LEU	HDB	0.769
143	LEU	HN	7.679
143	LEU	CA	51.777
143	LEU	CB	43.150
143	LEU	CG	26.919
143	LEU	CDA	25.656
143	LEU	CDB	23.130
143	LEU	N	119.448
144	PRO	HA	4.584
144	PRO	HBA	1.812
144	PRO	HBB	2.343
144	PRO	HGA	2.102
144	PRO	HGB	2.102
144	PRO	HDA	3.173
144	PRO	HDB	3.657
144	PRO	CA	62.170
144	PRO	CB	28.042
144	PRO	CG	27.752
144	PRO	CD	49.707
145	MET	HA	4.168
145	MET	HBA	2.192
145	MET	HBB	2.307
145	MET	HGA	1.440
145	MET	HGB	2.007
145	MET	HN	7.626
145	MET	CA	54.132
145	MET	CB	33.068
145	MET	CG	32.789
145	MET	N	118.081
146	HIS	HA	4.909
146	HIS	HBA	2.758
146	HIS	HBB	3.541

No.	Residue	Atom	Shift
146	HIS	HDA	7.491
146	HIS	HN	8.664
146	HIS	CA	55.762
146	HIS	CB	31.065
146	HIS	N	117.966
147	ILE	HA	4.584
147	ILE	HB	1.924
147	ILE	HGA	0.920
147	ILE	HG1A	1.681
147	ILE	HG1B	1.681
147	ILE	HD1	0.919
147	ILE	HN	7.661
147	ILE	CA	60.104
147	ILE	CB	40.150
147	ILE	CG1	29.220
147	ILE	CG2	14.489
147	ILE	CD1	16.968
147	ILE	N	117.232
148	ARG	HA	4.924
148	ARG	HBA	1.741
148	ARG	HBB	1.596
148	ARG	HGA	1.391
148	ARG	HGB	1.391
148	ARG	HDA	2.986
148	ARG	HDB	2.986
148	ARG	HN	8.280
148	ARG	CA	55.069
148	ARG	CB	33.071
148	ARG	CG	27.142
148	ARG	CD	43.478
148	ARG	N	126.523
149	LEU	HA	4.657
149	LEU	HBA	1.342
149	LEU	HBB	1.152
149	LEU	HG	2.865
149	LEU	HDA	0.768
149	LEU	HDB	0.819
149	LEU	HN	9.097
149	LEU	CA	53.683
149	LEU	CB	45.835
149	LEU	CG	26.667
149	LEU	CDA	23.525
149	LEU	CDB	19.556
149	LEU	N	123.699
150	SER	HA	4.782
150	SER	HBA	3.728
150	SER	HBB	3.728
150	SER	HG	5.324
150	SER	HN	8.147
150	SER	CA	57.321
150	SER	CB	65.478
150	SER	N	116.451

No.	Residue	Atom	Shift
151	PHE	HBA	2.720
151	PHE	HBB	3.124
151	PHE	HD1	7.375
151	PHE	HDA	7.375
151	PHE	HE1	7.250
151	PHE	HEA	7.250
151	PHE	HN	7.967
151	PHE	CA	57.078
151	PHE	CB	41.921
151	PHE	CD1	132.970
151	PHE	CD2	132.970
151	PHE	CE1	131.209
151	PHE	CE2	131.209
151	PHE	N	119.835
152	ASN	HA	4.942
152	ASN	HBA	2.562
152	ASN	HBB	2.768
152	ASN	HN	8.689
152	ASN	HD1A	6.928
152	ASN	HD1B	7.517
152	ASN	CA	50.904
152	ASN	CB	38.159
152	ASN	N	119.260
152	ASN	ND2	113.458
153	PRO	HA	3.877
153	PRO	HBA	1.852
153	PRO	HBB	2.316
153	PRO	HGA	1.810
153	PRO	HGB	2.207
153	PRO	HDA	3.789
153	PRO	HDB	3.789
153	PRO	CA	66.316
153	PRO	CB	32.185
153	PRO	CG	28.653
153	PRO	CD	50.594
154	THR	HA	3.832
154	THR	HB	4.090
154	THR	HG1	4.661
154	THR	HGA	1.112
154	THR	HN	7.719
154	THR	CA	65.756
154	THR	CB	68.679
154	THR	CG2	21.818
154	THR	N	111.383
155	GLN	HA	3.753
155	GLN	HGA	2.234
155	GLN	HGB	2.367
155	GLN	HN	7.723
155	GLN	HE1A	6.964
155	GLN	HE1B	7.559
155	GLN	CA	57.946
155	GLN	CB	29.238

No.	Residue	Atom	Shift
155	GLN	CG	34.854
155	GLN	N	119.140
155	GLN	NE2	111.805
156	LEU	HA	3.075
156	LEU	HBA	1.182
156	LEU	HBB	1.551
156	LEU	HG	1.629
156	LEU	HDA	0.605
156	LEU	HDB	0.605
156	LEU	HN	7.453
156	LEU	CA	57.725
156	LEU	CB	41.406
156	LEU	CG	27.497
156	LEU	CDA	25.111
156	LEU	CDB	25.111
156	LEU	N	115.674
157	GLU	HA	4.294
157	GLU	HGA	2.235
157	GLU	HGB	2.372
157	GLU	HN	7.044
157	GLU	CA	55.885
157	GLU	CG	33.007
157	GLU	N	112.453
158	GLU	HA	4.403
158	GLU	HBA	1.805
158	GLU	HBB	2.012
158	GLU	HN	7.132
158	GLU	N	118.631
159	GLN	HBA	1.988
159	GLN	HBB	1.988
159	GLN	HGA	2.396
159	GLN	HGB	2.517
159	GLN	HE1A	6.798
159	GLN	HE1B	7.474
159	GLN	NE2	111.965
160	CYS	HN	8.214
160	CYS	N	124.15

Appendix B

Semi-Automated Assignment Macros

```
from ccpnmr.analysis.ExperimentBasic import getSpectra
from ccpnmr.analysis.AssignmentBasic import makeResonanceGuiName, findMatchingShifts, assignResToDim,
getResonanceAtomTuple
from sast import areResonancesBound, areAtomsBound
from ccpnmr.analysis.AssignmentBasic import clearPeakDim, makeResonanceGuiName
from ccpnmr.analysis.PeakBasic import pickPeak, deletePeak
from ccpnmr.analysis.MarkBasic import createPeakMark, removeMarks
from C13_fullauto import isHeteroN, is HeteroCaro

def fullyauto(argServer=None, peakList=None):

  assert argServer or peakList



  if argServer:
    project =  argServer.getProject()

 #this bit gets all the spectra we're going to use

  spectra=getSpectra(project, minNumDim = 3)
  for experiment in project.nmrExperiments:

    print experiment.name
    if experiment.name=='15N_noesy':
      noesy=experiment
    elif experiment.name=='13C_noe_15NH':
      xtra=experiment
    elif experiment.name=='13C_noesy':
      Cnoesy=experiment


  quit=0


  #now we get the peaklists that we need

  if not peakList:
    for peakLista in noesy.dataSources[0].peakLists:
      if peakLista.serial==1:
        peakList=peakLista.peaks
    for peakListb in xtra.dataSources[0].peakLists:
      if peakListb.serial==1:
        xpeakList=peakListb.peaks
    for peakListc in Cnoesy.dataSources[0].peakLists:
      if peakListc.serial==1:
        CpeakList=peakListc


  #start with residue 1 and group all peaks with resnol and same restype together

  for n in range(127):            #number of residues
    if not quit:
      Hgroups={}
      Cshifts={}
      for peak in peakList:
        if peak.details:
          continue
        else:

          residue=peak.peakDims[0].peakDimContribs[0].resonance.resonanceSet.atomSets[0].atoms[0].residue
          restype=peak.peakDims[0].peakDimContribs[0].resonance.name
          if n==residue.seqCode:
            if Hgroups:
              for groupno in Hgroups.keys():
               if restype == Hgroups[groupno][0].peakDims[0].peakDimContribs[0].resonance.name:
                 Hgroups[groupno].append(peak)

                 break
              else:
                groupno+=1
                Hgroups[groupno]=[]
                Hgroups[groupno].append(peak)

            else:
              Hgroups[1]=[]
              Hgroups[1].append(peak)


  for Cpeak in xpeakList:
    if Cpeak.details:
      continue
    else:
     if Cpeak.peakDims[0].peakDimContribs:
       residue=Cpeak.peakDims[0].peakDimContribs[0].resonance.resonanceSet.atomSets[0].atoms[0].residue
       restype=Cpeak.peakDims[0].peakDimContribs[0].resonance.name
     else:
      continue
     if n==residue.seqCode:
       if Hgroups:
         for groupno in Hgroups.keys():
          if restype == Hgroups[groupno][0].peakDims[0].peakDimContribs[0].resonance.name:
            peakDim=Cpeak.peakDims[2]
            dataDimRef=peakDim.dataDimRef
            shift=peakDim.position
            if groupno in Cshifts.keys():

              Cshifts[groupno].append((findMatchingShifts(dataDimRef, shift, tolerance=0.3)))
              Cshifts[groupno][-1].append(Cpeak)
              break

            else:

              Cshifts[groupno]=[(findMatchingShifts(dataDimRef, shift, tolerance=0.3))]
              Cshifts[groupno][0].append(Cpeak)

              break

  for groupno in Hgroups.keys():
    if not quit:
      for peak in Hgroups[groupno]:
        popup=argServer.parent.getWindowPopup('window5')

        popup.gotoPeak(peak)
        peakDim=peak.peakDims[1]
        dataDimRef=peakDim.dataDimRef
        shift=peakDim.position
        position=[]
        position.append(peakDim.value)
        position.append(peak.peakDims[0].value)
        for shift in findMatchingShifts(dataDimRef, shift, tolerance=0.03):
          resonance=makeResonanceGuiName(shift.resonance)
          print 'this is', resonance
          if isHeteroN(shift.resonance):
           residue=shift.resonance.resonanceSet.atomSets[0].atoms[0].residue
           for resonance in residue.resonanceGroups[0].resonances:
             if areResonancesBound(shift.resonance, resonance) and
resonance.resonanceSet!=shift.resonance.resonanceSet:
               Nres=resonance

               position.append(Nres.shifts[0].value)
               newpeak=pickPeak(peakLista, position, unit='ppm')
               createPeakMark(newpeak)
               popup=argServer.parent.getWindowPopup('window2')   #window for Nnoesy
               popup.gotoPeak(newpeak)
               if argServer.askYesNo(message='Is there a symmetry related peak?'):
                 assignResToDim(peakDim, shift.resonance)
                 newpeak.delete()

               else:
                 newpeak.delete()
```

```
            del position[-1]


        elif isHeteroCaro(shift.resonance):
         for group in Cshifts[groupno]:

           for Cshift in group[:-1]:

             if areResonancesBound(shift.resonance, Cshift.resonance):
               position.append(Cshift.value)


               newpeak=pickPeak(peakListc, position, unit='ppm')
               createPeakMark(newpeak)
               popup=argServer.parent.getWindowPopup('window3')        #window for Cnoesy
               popup.gotoPeak(newpeak)
               if argServer.askYesNo(message='Is there a symmetry related peak?'):
                 assignResToDim(peakDim, shift.resonance)
                 newpeak.delete()
                 group[-1].setDetails('Done')
               else:
                 newpeak.delete()


               del position[-1]

        else:
         if shift.resonance.resonanceSet:
           residue=shift.resonance.resonanceSet.atomSets[0].atoms[0].residue
         else:
           continue
         for resonance in residue.resonanceGroups[0].resonances:
           if areResonancesBound(shift.resonance, resonance):
             if resonance.shifts and resonance.resonanceSet!=shift.resonance.resonanceSet:
               arores=resonance
               position.append(arores.shifts[0].value)
               newpeak=pickPeak(peakListc, position, unit='ppm')
               createPeakMark(newpeak)
               popup=argServer.parent.getWindowPopup('window3')   #window for Cnoesy
               popup.gotoPeak(newpeak)
               if argServer.askYesNo(message='Is there a symmetry related peak?'):
                 assignResToDim(peakDim, shift.resonance)
                 newpeak.delete()

               else:
                 newpeak.delete()


               del position[-1]




        removeMarks(project)

      peak.setDetails('Done')

    if argServer.askYesNo(message='Next strip?'):
      continue
    else:
      quit=1
      break
```

```
from ccp.api import Nmr
from ccpnmr.analysis.PeakBasic import findMatchingPeaks, aliasedPeakDimPosition, addPeakToSelected,
removePeakFromSelected
from memops.gui.Frame import Frame
from memops.gui.Label import Label
from memops.gui.LabelFrame import LabelFrame
from memops.gui.PulldownMenu import PulldownMenu
from memops.gui.ScrolledMatrix import ScrolledMatrix
from memops.gui.Util import createDismissHelpButtonList
from memops.gui.MessageReporter import showError
from ccpnmr.analysis.UnitConverter import unit_converter,pnt2ppm
from ccpnmr.analysis.Util import setDataDimAssignmentTolerance, getDataDimAssignmentTolerance
from ccpnmr.analysis.AssignmentBasic import makeResonanceGuiName, findMatchingShifts, assignResToDim,
getResonanceAtomTuple
import string
import difflib
from Tkinter import *
import sys
import autoass
from sast import symfinder, CHlister

def philtr(argServer=None, peaks=None):


        assert argServer or peaks

        class possible:
          def __init__(self, shift, peak, peakDim, counter):
            self.name=makeResonanceGuiName(shift.resonance)
            self.delta=abs(shift.value-peakDim.value)
            self.peak=counter
            self.intensity=peak.peakIntensities[0].value
            self.Hres=shift.resonance
            self.atomtype=list(getResonanceAtomTuple(shift.resonance)[-1])
            self.shift=shift.value
            if shift.resonance.resonanceSet:
              if isHeteroN(shift.resonance):
                self.X='N'
              else:
                self.X='C'
        class valid:
          def __init__(self, H,X):
            self.Hname=H.name
            self.Xname=X.name
            self.Hdelta=H.delta
            self.Xdelta=X.delta
            self.Hpk=H.peak
            self.Xpk=X.peak
            self.Hres=H.Hres
            self.Xres=X.Hres
             self.Xshift=X.shift
            self.Hintensity=H.intensity
            self.Xintensity=X.intensity
             self.X=H.X



        Cpeaks=[]
        Hpeaks=[]
        Npeaks=[]
        Cobjects=[]
        CHobjects=[]
        NHobjects=[]
        Nobjects=[]
        validpeaks={}
        Cintensities=[]
        Nintensities=[]
        NHintensities=[]
```

```
CHintensities=[]
reducedambiguity={}
totalambiguity={}
aro=['PHE','TYR','TRP','HIS']
aro2=['H','HA', 'HBa', 'HBb', 'HB*']

def areAtomsBound(atom1, atom2):

   for chemBond in atom1.chemAtom.chemBonds:
     for chematom in chemBond.chemAtoms:
       if chematom.name==atom2.name:
         return 1
   return 0

 def areResonancesBound(resonance1,resonance2):
   if resonance1.resonanceSet and resonance2.resonanceSet:
     residue1 = resonance1.resonanceSet.atomSets[0].atoms[0].residue
     residue2 = resonance2.resonanceSet.atomSets[0].atoms[0].residue
   else:
    residue1=None
    residue2=1
   # quickly discard on the basis of residue
   if residue1 is not residue2:
     return 0

   for atomSet1 in resonance1.resonanceSet.atomSets:
     for atom1 in atomSet1.atoms:
       for atomSet2 in resonance2.resonanceSet.atomSets:
         for atom2 in atomSet2.atoms:
          if areAtomsBound(atom1,atom2):
            return 1
   return 0


def isHeteroN(resonance):
  if resonance.resonanceSet:
    for atomSet in resonance.resonanceSet.atomSets:
      for atom in atomSet.atoms:
        for chemBond in atom.chemAtom.chemBonds:
          for chematom in chemBond.chemAtoms:
            if 'N' in list(chematom.name):
              return 1
   return 0


if peaks is None:
  peaks = argServer.getCurrentPeaks()

for peak in peaks:

  if peak.peakList.serial==2 and peak.peakList.dataSource.experiment.name=='15Nnoe':
    Hpeaks.append(peak)
  elif peak.peakList.dataSource.experiment.name=='C_N_noe':
    Cpeaks.append(peak)
  elif peak.peakList.dataSource.experiment.name=='15N_noe_15N':
    Npeaks.append(peak)


isotope=peaks[0].peakDims[-1].dataDim.dataDimRefs[0].expDimRef.isotopeCodes

project = argServer.getProject()

Ccounter=1
for peak in Cpeaks:
  resname=[]
  peakDim=peak.peakDims[-1]
  dataDimRef=peakDim.dataDimRef
  dataDim=dataDimRef.dataDim

  position=peakDim.position
  Cintensities.append(peak.peakIntensities[0].value)
```

```
    for shift in findMatchingShifts(dataDimRef, position, tolerance=0.27):

      if shift.resonance.resonanceSet:
        Cpkobject=possible(shift, peak, peakDim, Ccounter)
        Cobjects.append(Cpkobject)

    Ccounter=Ccounter+1


  Cratios=[]
  Ccopy=Cintensities[:]
  Ccopy.sort()
  if Cintensities:
    for Cint in Cintensities:
      Cratios.append(Cint/Ccopy[-1])



  Ncounter=1
  for peak in Npeaks:
    peakDim=peak.peakDims[-2]
    dataDimRef=peakDim.dataDimRef
    dataDim=dataDimRef.dataDim
    position=peakDim.position + peakDim.numAliasing*dataDim.numPoints
    Nintensities.append(abs(peak.peakIntensities[0].value))


    for shift in findMatchingShifts(dataDimRef, position, tolerance=0.19):
      print Ncounter, makeResonanceGuiName(shift.resonance), shift
      Npkobject=possible(shift, peak, peakDim, Ncounter)
      Nobjects.append(Npkobject)
    Ncounter+=1



  Nratios=[]
  Ncopy=Nintensities[:]
  Ncopy.sort()
  if Nintensities:
    for Nint in Nintensities:
      Nratios.append(Nint/Ncopy[-1])


  print len(Hpeaks)


  Hcounter=1

  for Hpeak in Hpeaks:

    allHpeaks={}
    allHpeakname=[]
    diagnosis=[]
    peakDim=Hpeak.peakDims[-2]
    position=peakDim.position
    dataDimRef=peakDim.dataDimRef
    dataDim=peakDim.dataDimRef.dataDim
    isotope=peakDim.dataDim.dataDimRefs[0].expDimRef.isotopeCodes



    totalambiguity[Hcounter]=0
    if findMatchingShifts(dataDimRef, position, tolerance=0.032):
      for shift in findMatchingShifts(dataDimRef, position, tolerance=0.032):
        totalambiguity[Hcounter]+=1
        if shift.resonance.resonanceGroup:
          if str(shift.resonance.resonanceGroup.residue.ccpCode) in aro and str(shift.resonance.name) not in
aro2:
            if Hcounter in validpeaks.keys():
```

```
              validpeaks[Hcounter].append(shift)
              Hpkobject=None
            else:
              validpeaks[Hcounter]=[shift]
              Hpkobject=None
          else:
            Hpkobject=possible(shift, Hpeak, peakDim, Hcounter)


        else:          #if not Hpkobject.resonance.resonanceGroup:
          Hpkobject=None    # pass


        if Hpkobject:
          if isHeteroN(Hpkobject.Hres):
            NHobjects.append(Hpkobject)
            if Hpeak.peakIntensities[0].value not in NHintensities:
              NHintensities.append(abs(Hpeak.peakIntensities[0].value))
              CHintensities.append(0)
          else:
            CHobjects.append(Hpkobject)

            if Hpeak.peakIntensities[0].value not in CHintensities:

              CHintensities.append(abs(Hpeak.peakIntensities[0].value))
              NHintensities.append(0)
    else:
      validpeaks[Hcounter]=[]
      totalambiguity[Hcounter]=0
      NHintensities.append(0)
      CHintensities.append(0)
    Hcounter=Hcounter+1



  NHratios=[]
  if NHintensities:
    NHcopy=NHintensities[:]
    NHcopy.sort()
    print NHcopy
    if NHcopy[-1]!=0:
      for NHint in NHintensities:
        NHratios.append(NHint/NHcopy[-1])

  CHratios=[]
  if CHintensities:
    CHcopy=CHintensities[:]
    CHcopy.sort()
    if CHcopy[-1]!=0:
      for CHint in CHintensities:
        CHratios.append(CHint/CHcopy[-1])



  for H in CHobjects:
    for C in Cobjects:
      if areResonancesBound(H.Hres, C.Hres):
        validpk=valid(H,C)
        if validpk.Hpk not in validpeaks.keys():
          validpeaks[validpk.Hpk]=[validpk]
        else:
          validpeaks[validpk.Hpk].append(validpk)
      elif H.peak not in validpeaks.keys():
        validpeaks[H.peak]=[]
  for H in NHobjects:
    for N in Nobjects:
      if areResonancesBound(H.Hres, N.Hres):
```

```
      validpk=valid(H,N)
      if validpk.Hpk in validpeaks.keys():
        validpeaks[validpk.Hpk].append(validpk)
      else:
        validpeaks[validpk.Hpk]=[validpk]
       print validpk.Hpk, validpk.Xres


    elif H.peak not in validpeaks.keys():
      validpeaks[H.peak]=[]

print validpeaks


class interface(Frame):
  def __init__(self, root, validpeaks, Hpeaks, Cratio, Nratios, CHratios, NHratios,  totalambiguity, peaks):
    Frame.__init__(self, root)
    self.root=root
    self.pks=validpeaks
    self.peaks=peaks
    self.keys=self.pks.keys()
    self.Hpeaks=Hpeaks
    self.resonance={}
    self.states=[]
    self.usedNpeaks=[]
    self.usedCpeaks=[]
    self.partusedNpeaks=[]
    self.partusedCpeaks=[]
    self.display()
    self.matches=[]
    self.Cratios=Cratios
    self.Nratios=Nratios
    self.CHratios=CHratios
    self.NHratios=NHratios
    self.totalambiguity=totalambiguity



  def display(self):#, Hpeaks):
    Label(self.root, text='Select a peak', bg='red').grid(sticky=NSEW)
    box=Listbox(self.root, relief='sunken')
    sbar=Scrollbar(self.root)
    sbar.config(command=box.yview)
    box.config(yscrollcommand=sbar.set)
    sbar.grid(row=1, column=1, sticky=NSEW)
    box.grid(row=1, column=0)
    for peak in self.pks.keys():
      box.insert(END, peak)
    box.config(selectmode=SINGLE, setgrid=1)
    self.box=box
    #self.Hpeak=self.Hpeaks[0]
    Button(root, text='Assign', command=self.onPoss).grid()
    Button(root, text='Check', command=self.onChk).grid()
    Button(root, text='Assign other dims', command=self.onallassign).grid()


  def onChk(self):
    unusedpeaks={}
    print 'check in progress'
    for key in self.keys:
      for pk in self.pks[key]:
        if hasattr(pk, 'Xpk'):
          if pk.X=='C' and pk.Xpk in self.usedCpeaks:
            pass
          elif pk.X=='C' and pk.Xpk in self.partusedCpeaks:
            pass
          elif pk.X=='C' and pk.Xpk not in unusedpeaks.keys():
            unusedpeaks[pk.Xpk]=pk
    if unusedpeaks:
      print len(unusedpeaks)
      print 'unused peaks found'
      win=Toplevel()
      Label(win, text='Unused Peaks', bg='red').grid(sticky=NSEW)
      r=2
      for pk in unusedpeaks.values():

        Label(win, text='%s, shift = %g' %(pk.Xpk, pk.Xshift), relief=GROOVE).grid(row=r, sticky=NSEW)
        r+=1
    unusedpeaks={}

  def onPoss(self):

    for pk in self.peaks:
      if pk in argServer.getCurrentPeaks():
        peaks=self.peaks
        continue
      else:
        peaks=argServer.getCurrentPeaks()


    self.number=int(self.box.get(ACTIVE))
    self.index=int(self.box.get(ACTIVE))-1
    self.Hpeak=self.Hpeaks[self.index]
    if symfinder(self.Hpeak, Hpossibles=self.pks[self.number], pks=peaks):
      self.key=symfinder(self.Hpeak, Hpossibles=self.pks[self.number], pks=peaks)
    else:
      self.key=self.pks[self.number]
    self.states=[]
    print self.Hpeak.peakDims[-2].value

    if self.key!=[]:


      for i in range(len(self.key)):


        if hasattr(self.key[i], 'X') and self.key[i].X=='N':
          self.win=Toplevel()
          self.Titles=['Hposs', 'Hdelta' ,'Npk',  'Nposs','Nshift','Ndelta', 'Intensity']
          Label(self.win, text='Possibilities for Hpk %d, shift=%g, intensity=%g' %(self.number,
self.Hpeak.peakDims[-2].value, self.NHratios[self.index]), bg='red').grid(columnspan=8, sticky=NSEW)
          self.window(peaks=self.key, ratios=Nratios,usedpks=self.usedNpeaks ,
partusedpks=self.partusedNpeaks )
          break
        elif hasattr(self.key[i], 'X') and self.key[i].X=='C':
          self.win=Toplevel()
          self.Titles=['Hposs', 'Hdelta',  'Cpk',  'Cposs','Cshift','Cdelta','Intensity']
          Label(self.win, text='Possibilities for Hpk %d, shift=%g, intensity=%g' %(self.number,
self.Hpeak.peakDims[-2].value, self.CHratios[self.index]), bg='red').grid(columnspan=8, sticky=NSEW)
          self.window(peaks=self.key, ratios=Cratios, usedpks=self.usedCpeaks ,
partusedpks=self.partusedCpeaks)
          break
        elif i==len(self.key)-1:
          self.win=Toplevel()
          self.Titles=['Hposs', 'Hdelta',  'Cpk',  'Cposs','Cshift','Cdelta','Intensity']
          self.window(peaks=self.key)
          break
      else:
        win=Toplevel()
        if self.NHratios[self.index]!=0.0 and self.CHratios:
          Label(win, text='Possibilities for Hpk %d, shift=%g, intensity=%g(NH) or %g(CH)'
%(self.number,self.Hpeak.peakDims[-2].value, self.NHratios[self.index], self.CHratios[self.index]),
bg='red').grid(columnspan=5, sticky=NSEW)
          Label(win, text='No matches' , bg='red').grid(columnspan=5, sticky=NSEW)
        elif self.NHratios[self.index]!=0.0 and not self.CHratios:
          Label(win, text='Possibilities for Hpk %d, shift=%g, intensity=%g(NH)'
%(self.number,self.Hpeak.peakDims[-2].value, self.NHratios[self.index]), bg='red').grid(columnspan=5, sticky=NSEW)
          Label(win, text='No matches' , bg='red').grid(columnspan=5, sticky=NSEW)
        else:
          Label(win, text='Possibilities for Hpk %d, shift=%g, intensity=%g'
%(self.number,self.Hpeak.peakDims[-2].value, self.CHratios[self.index]), bg='red').grid(columnspan=5, sticky=NSEW)
          Label(win, text='No matches' , bg='red').grid(columnspan=5, sticky=NSEW)
```

```
def window(self, peaks=None, ratios=None, usedpks=None, partusedpks=None):


 def onAssign():
    resonances=[]
    peakDim=self.Hpeak.peakDims[-2]

    for i in range(len(self.states)):

      if self.states[i].get():
         resonances.append(self.poss[i+2])
     self.states=[]
     if len(resonances)==1:
      if hasattr(resonances[0],'resonance'):
        resonance=resonances[0].resonance
      elif hasattr(resonances[0],'Hres'):
        resonance=resonances[0].Hres

      assignResToDim(peakDim, resonance)
      self.Hpeak.setDetails(str(totalambiguity[self.number])+' '+str(1))
      if hasattr(resonances[0],'X') and resonances[0].X=='N':
        self.usedNpeaks.append(resonances[0].Xpk)

       elif hasattr(resonances[0],'X') and resonances[0].X=='C':
        self.usedCpeaks.append(resonances[0].Xpk)
      self.poss={}

    else:
      self.Hpeak.setDetails(str(totalambiguity[self.number])+' '+str(len(resonances)))
      for item in resonances:
       if hasattr(item,'resonance'):
         res=item.resonance
       elif hasattr(item,'Hres'):
         res=item.Hres


       assignResToDim(peakDim, res)
       self.Hpeak.setDetails(self.Hpeak.details+' '+str(makeResonanceGuiName(res)))
       if hasattr(resonances[0],'X') and resonances[0].X=='N':
         self.partusedNpeaks.append(item.Xpk)
       elif hasattr(resonances[0],'X') and resonances[0].X=='C':
         self.partusedCpeaks.append(item.Xpk)
       self.poss={}


def selectAll():
  print self.states
  for variable in self.states:
    print variable
    variable.set(1)
    print variable.get()


def onAssignHxab():
  resonances=[]
  peakDim=self.Hpeak.peakDims[-2]

  for i in range(len(self.states)):
    if self.states[i].get():
       resonances.append(self.poss[i+2])
  for item in resonances:
    if hasattr(item,'resonance'):
     res=item.resonance
    elif hasattr(item,'Hres'):
     res=item.Hres
    assignResToDim(peakDim, res)
    self.Hpeak.setDetails(str(totalambiguity[self.number])+' '+str(1))
   if hasattr(resonances[0],'X') and resonances[0].X=='N':
    self.usedNpeaks.append(item.Xpk)
   elif hasattr(resonances[0],'X') and resonances[0].X=='C':
    self.usedCpeaks.append(item.Xpk)

  col=0
  for Title in self.Titles:
    Label(self.win, text=Title, bg='white', relief=RIDGE).grid(row=1,column=col, sticky=NSEW)
    col+=1

  r=2
  self.poss={}
  for poss in peaks:

    var=IntVar()
    if hasattr(poss, 'Hname'):
      Label(self.win, text=poss.Hname, relief=GROOVE).grid(row=r,column=0, sticky=NSEW)
      Label(self.win, text=poss.Hdelta, relief=GROOVE).grid(row=r,column=1, sticky=NSEW)
      if poss.Xpk in usedpks:
        Label(self.win, text=poss.Xpk, bg='red', relief=GROOVE).grid(row=r,column=2, sticky=NSEW)
      elif poss.Xpk in partusedpks:
       Label(self.win, text=poss.Xpk, bg='yellow', relief=GROOVE).grid(row=r,column=2, sticky=NSEW)
     else:
       Label(self.win, text=poss.Xpk, relief=GROOVE).grid(row=r,column=2, sticky=NSEW)
     Label(self.win, text=poss.Xname, relief=GROOVE).grid(row=r,column=3, sticky=NSEW)
      Label(self.win, text=poss.Xshift, relief=GROOVE).grid(row=r,column=4, sticky=NSEW)
     Label(self.win, text=poss.Xdelta, relief=GROOVE).grid(row=r,column=5, sticky=NSEW)
      Label(self.win, text=ratios[int(poss.Xpk)-1], relief=GROOVE).grid(row=r,column=6, sticky=NSEW)
      self.win.rowconfigure(r, weight=1)
     self.poss[r]=poss
      self.states.append(var)
      Checkbutton(self.win, text=r-2, command=None, variable=var).grid(row=r, column=7, sticky=NSEW)
      r+=1
    else:
      self.poss[r]=poss
      self.states.append(var)
      Label(self.win, text=makeResonanceGuiName(poss.resonance)+' : aromatic possibility', bg='red',
relief=GROOVE).grid(row=r,columnspan=7, sticky=NSEW)
      Checkbutton(self.win, text=r-2, command=None, variable=var).grid(row=r, column=7, sticky=NSEW)
     r+=1

  for z in range(8):
    self.win.columnconfigure(z, weight=1)


 Button(self.win, text='Assign', command=onAssign).grid(row=r+1, column=1, sticky=NSEW)
 Button(self.win, text='Select all', command=selectAll).grid(row=r+1, column=2, sticky=NSEW)
Button(self.win, text='Assign Hx* Unambiguous', command=onAssignHxab).grid(row=r+1, column=3, sticky=NSEW)
 self.win.rowconfigure(r+1, weight=1)


def allAssign(self):
  for i in range(len(self.vars)):
    if self.vars[i].get():
     self.matches.append(self.matchedposs[i+1])

  if len(self.matches)==1:
    match=self.matches[0]
    for peak in self.Hpeaks:
      res1=match[0].resonance
      peakDim1=peak.peakDims[0]
      print res1.name
      print peakDim1.dataDim.dataDimRefs[0].expDimRef.isotopeCodes
      res2=match[-1].resonance
      peakDim2=peak.peakDims[-1]
      assignResToDim(peakDim1, res1)
```

```
      assignResToDim(peakDim2, res2)
      self.matches=[]
    else:
      showError('Error','Only one may be selected')
      self.matches=[]


def onallassign(self):
  win=Toplevel()

  shifts=[]
  Hpossibles=[]
  Npossibles=[]
  self.matchedposs={}
  L=[]
  self.vars=[]



  class possible:
    def __init__(self, shift, peakDim):
      self.peakDim=peakDim
      self.resonance=shift.resonance
      self.name=makeResonanceGuiName(shift.resonance)
      self.shift=shift.value
       self.residue=shift.resonance.resonanceSet.atomSets[0].atoms[0].residue

  for peak in self.Hpeaks:

   for peakDim in peak.peakDims:
     dataDimRef=peakDim.dataDimRef
     dataDim=dataDimRef.dataDim
     position=peakDim.position
     if peakDim.dataDim.dataDimRefs[0].expDimRef.isotopeCodes==('1H',):
       #tolerance=setDataDimAssignmentTolerance(dataDim, 0.035)
      for shift in findMatchingShifts(dataDimRef, position, tolerance=0.035):
        if shift.resonance.resonanceSet:

          if shifts:
            if shift.resonance in shifts:

              val=possible(shift, peakDim)
              Hpossibles.append(val)
            else:
              shifts.append(shift.resonance)
          else:
            shifts.append(shift.resonance)
     else:
       #tolerance=setDataDimAssignmentTolerance(dataDim, 0.1)
       for shift in findMatchingShifts(dataDimRef, position, tolerance=0.2):

        if shift.resonance:
          print shift.resonance.name
          if shifts:
            if shift.resonance in shifts:

              val=possible(shift, peakDim)
              Npossibles.append(val)
            else:
              shifts.append(shift.resonance)
          else:
            shifts.append(shift.resonance)
  c=1
  for item in Hpossibles:
    for other in Npossibles:
      match=[item, other]


      if areResonancesBound(item.resonance,other.resonance) and item.resonance not in L:
        L.append(item.resonance)
        self.matchedposs[c]=match
        c+=1


    if self.matchedposs:
      Label(win, text='Possibilities', bg='red').grid(columnspan=5,sticky=NSEW)
      Titles=['Name', 'Shift', 'Name','Shift']
      col=0
      for Title in Titles:
        Label(win, text=Title, bg='white', relief=RIDGE).grid(row=1,column=col, sticky=NSEW)
        col+=1

      r=2

      for key in self.matchedposs.keys():


        var=IntVar()
        atom1, atom2=self.matchedposs[key]
        Label(win, text=atom1.name, relief=GROOVE).grid(row=r,column=0, sticky=NSEW)
        Label(win, text=atom1.shift, relief=GROOVE).grid(row=r,column=1, sticky=NSEW)
        Label(win, text=atom2.name, relief=GROOVE).grid(row=r,column=2, sticky=NSEW)
        Label(win, text=atom2.shift, relief=GROOVE).grid(row=r,column=3, sticky=NSEW)
        Checkbutton(win, command=None, variable=var).grid(row=r, column=4, sticky=NSEW)
        self.vars.append(var)
        win.rowconfigure(r, weight=1)

        r+=1
      for z in range(5):
        win.columnconfigure(z, weight=1)
      Button(win, text='Assign', command=self.allAssign).grid()



root=Toplevel()
if validpeaks:
  top=interface(root, validpeaks, Hpeaks, Cratios, Nratios, CHratios, NHratios, totalambiguity, peaks)
  top.mainloop()
```